

Wanted: a free rein for Germany's universities

Against a background of progressive change in the running of Germany's research, the universities stand out as bastions of down-at-heel conservatism. Increased competition is the way forward.

At the start of this decade, prospects for German science were looking bleak. A variety of factors conspired to keep German laboratories closed to new and important research areas, particularly genetics and informatics, and to keep young blood yoked until middle age within overly rigid disciplinary boundaries and outdated systems of employment.

But in the past few years the landscape has begun to be transformed, thanks not least to a new generation of strong leaders in key positions. Former research minister Jürgen Rüttgers was the first to take a decisive stand. He relaxed Germany's notoriously bureaucratic laws regulating research in genetic manipulation, and in 1996 launched an astonishingly successful regional biotechnology competition, BioRegio, which cunningly forced the *Länder* (states) to face up to new economic opportunities. He created well-funded programmes for genome research and information sciences. Biotechnology companies sprang up in what had been a desert.

Hard on Rüttgers' heels came Hubert Markl and Ernst-Ludwig Winnacker, respective heads of the Max Planck Society and the Deutsche Forschungsgemeinschaft, the university granting agency, with a series of bold and important reforms; to name a few: the promotion of the academic careers of young scientists and women, the creation of graduate colleges to be run by universities and Max Planck institutes, and the concentration of money in strategic areas of basic research identified by the scientific community itself.

The universities have proven the last bastion against change. But now it seems that even they are starting to shift. The moves by the University Rectors' Conference (HRK) to recommend the creation of a type of short-term assistant professorship, and to push for performance-related pay, are significant (see page 299). In the new reform-oriented climate, no one doubts that the necessary changes to federal

law will be carried through. But equally important is investment in university infrastructure. Germany adheres with pride to the Humboldt principle of teaching in an environment of research, but whereas a Max Planck institute typically boasts world-class facilities, a neighbouring university laboratory is likely to look shabby by comparison.

The Max Planck Society is a strong umbrella organization; the universities, proudly independent, are weakened as a lobby group because of their individual isolation; that partly explains the decline in their funds. Moreover, independence is a relative term. Universities are financed by the *Länder* governments, and so are free from federal interference. But most universities have only recently been given the right to administer their own budgets (though few have as yet rushed to exploit the advantages). And only last year were they allowed to select a small proportion — 20 per cent — of their students.

To supplement the new reforms, universities must become much more independent. They must grasp the opportunities to establish competition for resources between faculties and research groups. But competition must also be introduced at the federal level. The law blocking universities from selecting their own students, built on the laudable but fiscally insupportable concept that all those with a high-school *Arbitur* should have the right to a university education, must be changed. Universities cannot go on sagging under the resource-draining weight of thousands of students they do not want to teach, a high proportion of whom will not complete their degree.

In under-subscribed but important courses like chemistry and physics, universities should be able to fight for the best. And the best students will want to attend a strong research university with a good teaching record. *Länder* governments will then need to worry about losing out to their neighbours and rivals, and be more ready to provide the necessary levels of core funding for their universities. □

Farewell to a friend

George Brown was science's best friend in Congress because he saw it as the only foundation for a just society.

In an age of growing public cynicism about politics, the United States Congress can ill afford to lose the likes of George Brown (Democrat, California), the senior minority member of the Science Committee of the House of Representatives, who died last week after serving in the Congress for 36 years.

Science in the United States is indebted to Brown not just for his support of scientific programmes — which was unwavering, if never uncritical — but also for his desire to ensure that scientific issues were aired with an eloquence and sophistication rarely found elsewhere on Capitol Hill.

Even as chair of the Science Committee, before 1995, Brown had limited influence in Washington. Like its approximate counterparts in many other countries, the Science Committee is a minor player in the Congress's power structure. As a lifelong liberal and pacifist, whose seniority had been curtailed when he stepped out of the House to

unsuccessfully run for a California Senate seat in 1970, Brown had no illusions about his own place within that structure.

But he showed that, even if its budgetary influence was small, it was possible for the committee to serve as an effective forum for scientific ideas. With the aid of a strong specialist staff (rather than the usual collection of political hangers-on), Brown managed to nurture a discussion of scientific subjects in the US Congress that transcended 'pork barrel' politics and helped to create an environment in which science could better support good government.

Brown's motive for this endeavour was never doubted, and was indeed invoked last Friday in a moving tribute from his friend and rival, Jim Sensenbrenner (Republican, Wisconsin), the current chair of the Science Committee. Quite simply, Brown believed that science could form the basis of a more just and peaceful world. Researchers can pay tribute to his memory by keeping that noble aim in mind. □



US researchers losing battle to hide their data from prying eyes

RICHARD T. NOVITZ

[WASHINGTON] Scientists' representatives are losing a fight to protect US researchers from rules which, they say, could wreak havoc in laboratories by allowing outside parties to examine federally funded data at any time — perhaps even before publication — under freedom of information laws.

Defeat in a crucial vote of the House Appropriations Committee, followed by the death on 15 July of George Brown (Democrat, California), the senior Democrat on the House Science Committee and a vigorous opponent of the rules, leaves the way open for them to go into force in October, say science lobbyists.

Bruce Alberts, the president of the National Academy of Sciences, and Harold Varmus, director of the National Institutes of Health, both testified in Congress last week that the proposed rules would damage the scientific process and open researchers in controversial fields to endless harassment by special-interest groups.

"There are pitfalls in unrestrained openness," Varmus warned. These included "unwarranted violations of privacy, the



Alberts: new rules "invite harassment".

potential harassment of scientific investigators, and the chilling effect that such scrutiny could have on the free exchange of ideas".

The rules were initially proposed by Senator Richard Shelby (Republican, Alabama), and quietly passed into law last October as part of a huge budget bill. They would allow interested parties to request data from all government-funded scientists by issuing a request under the Freedom of Information Act (FOIA), the legislation intended to prevent the federal government keeping secrets from the public.

The law requires the White House Office of Management and Budget (OMB) to draw up rules for the application of FOIA to university research (see *Nature* 397, 459; 1999). But many believe that the Shelby provision is

primarily intended to let industry groups challenge university scientists researching the effects of pollution, tobacco or firearms.

"This actually invites harassment of scientists by those who don't like what the scientist is doing," Alberts told a 15 July hearing of the House subcommittee on government management, information and technology.

Pro-business groups and a former Republican OMB director testified against legislation proposed by Brown that would have overturned the Shelby provision. Alberts and Varmus supported the Brown legislation. But Brown, who had been ill since contracting an infection during heart surgery in May, was unable to testify, and died shortly after the hearing (see pages 295 and 305).

Although Stephen Horn (Republican, California), chairman of the government management subcommittee, was non-committal, the rest of his subcommittee divided along party lines, with the Republican majority opposing the Brown legislation.

Two days earlier, scientists' efforts to block the Shelby regulation had suffered a serious setback, when the powerful Appropriations Committee voted 33–25 against an amendment put forward by James Walsh (Republican, New York) and David Price (Democrat, North Carolina) to suspend its implementation for a year.

Science lobbyists say that the amendment fell after intense lobbying by groups including the National Rifle Association and the US Chamber of Commerce, which sees Shelby's seemingly arcane proposal as a key means of rolling back government regulation.

According to a circular issued by the chamber in March, the regulation "will do more for regulatory reform than all the legislation passed in the last ten years" if it is "implemented properly". Industry groups believe that the rule will enable them to challenge scientific studies that agencies, such as the Environmental Protection Agency, use to justify new regulations.

The OMB is seeking to lessen the impact of the provision when it publishes the detailed rule under which it will be implemented. In a second draft, due to be released for public comment shortly, it seeks to introduce a definition of scientific data that will exclude personal medical data and commercial proprietary information.

But opponents of the regulation say that such distinctions may not stand up in court, where the Freedom of Information Act could be interpreted as requiring all information to be released. "The Shelby amendment is fatally flawed," says Alberts.

Colin Macilwain

Computer upgrade for Canadian science

[MONTREAL] Canadian scientists keen to gain access to high-performance computing facilities will benefit from an award of Can\$23 million (US\$15.5 million) from the Canada Foundation for Innovation (CFI).

The new money, combined with expected contributions from other sources, means that new research facilities worth more than Can\$70 million will be available for research in areas as diverse as climate prediction, engineering, industrial design, medicine, astronomy and chemistry.

Canada's CaNet 3 facilities, offering state-of-the-art Internet access, will make possible a grid of computational resources that will encourage close collaboration between research institutions across Canada. That should help to "preserve one of our country's most important resources, its intellectual capital," says David Strangway, CFI's president and chief executive officer.

The awards for high-performance computing facilities were among a total investment by the CFI of Can\$226 million to support 122 infrastructure projects in 51 universities, colleges, hospitals and non-profit research institutions.

According to Strangway, the money will be joined by an additional Can\$339 million from partners in the public, private and

voluntary sectors. Since August 1998, awards by the CFI, a non-profit corporation established by the federal government in 1997, have totalled Can\$426 million. Contributions from other partners have resulted in a total capital investment of more than Can\$1 billion in research facilities.

The national network of shared computational research facilities will be created by a new Canadian initiative called C3.ca, an umbrella organization set up in March. Brian Unger, president of C3.ca, says CFI's support "towards a world-class infrastructure will enable us to more successfully compete at an international level".

By using parallel architecture, the network will allow some of the largest computers in the world to operate in unison for research not only in science and engineering, but also in finance, business and the arts.

Six universities and regional consortia, all members of C3.ca, were approved for CFI funding. They will establish facilities for computation and visualization, including about a dozen parallel shared memory and vector systems for advanced computing, along with six new multimedia visualization centres.

David Spurgeon

US sends mixed message in GM debate...

[WASHINGTON] In what had been billed as a major policy speech, US Agriculture Secretary Dan Glickman moved to ease public worries over genetically modified (GM) crops last week by acknowledging the need for unbiased research that would assure their safety.

But he stopped short of calling for the compulsory labelling of GM foods, and spoke of a US willingness to "vigorously fight for our legitimate rights" in trade negotiations with Europe over biotechnology products.

In his speech, Glickman announced several steps designed to bolster consumer confidence in GM foods. The US Department of Agriculture (USDA), which is responsible for evaluating the risks of biotechnology products to plants and animals, will initiate an independent review of the science behind product approval. This would presumably be conducted by the National Research Council, part of the National Academy of Science complex, which is already examining US Environmental Protection Agency (EPA) regulation of GM pest-resistant plants.

Glickman stressed the importance of separating USDA's research activities from its efforts to promote US agriculture "to ensure that no commercial interests have even the appearance of influence on our decisions regarding food safety". He also called for the establishment of regional centres "to evaluate biotech products over a long period of time and to provide information on an ongoing basis to growers, consumers, researchers and regulators".

But details of the work these centres will do remains sketchy, and it is uncertain whether they will conduct long-term ecological studies of GM crops. USDA spokesman



Field work: Glickman (centre, in tie) is seeking to satisfy both advocates and critics of GM crops.

Andy Solomon said there will be 8 to 12 such centres, which would be funded next year from the department's existing budget, with a request for additional funding likely in 2001.

Just as vague was Glickman's call for "all developers of biotech products to report any unexpected or potentially adverse effects to the Department of Agriculture immediately upon discovery". Val Giddings, vice-president for food and agriculture at the Biotechnology Industry Organization (BIO), points out that this requirement already exists for GM foods undergoing field tests before commercial release. Any reporting after commercial release would probably be strictly voluntary, says a USDA source.

Glickman received mostly favourable reactions from both sides of the GM food debate. BIO president Carl Feldbaum called it a "good speech" and said "I don't think it was in any way a blow" to the biotech industry.

Rebecca Goldburg of the Environmental Defense Fund, which has led the campaign to raise awareness of the risks of GM crops, said that Glickman "lent considerable legitimacy" to concerns about those risks.

Glickman acknowledged reports of possible harm to monarch butterflies (see *Nature* 399, 287; 1999) but added that there has been no evidence of any harmful effects in the field.

"I believe that distrust [of GM foods] is scientifically unfounded," he said. European reaction "comes in part from the lack of faith in the European Union to assure the safety of their food. They have no independent regulatory agencies like the [Food and Drug Administration], USDA or EPA."

Regarding the dispute over whether GM foods should be labelled in supermarkets — a practice the US biotech industry opposes — Glickman was again diplomatic. "At the end of the day, many observers, including me, believe some type of informational labelling is likely to happen," he said. "But I do believe that it is imperative that such labelling does not undermine trade and this promising new technology."

On the matter of trade negotiations with Europe, Glickman was uncompromising and emphatic. "We cannot let others hide behind unfounded, unwarranted scientific claims to block commerce in agriculture," he said. This, and his statements espousing the promise of GM crops, led Giddings to pronounce the speech a "robust endorsement of biotechnology".

Yet Glickman also called for the biotech industry to act responsibly and take public concern into account. "Product development to date has enabled those who oppose this technology to claim that all the talk about feeding the world is simply cover for corporate profit-making," he said, adding that "industry needs to act with greater sensitivity and foresight".

Tony Reichhardt

... as questions emerge over cost-effectiveness

[WASHINGTON] The jury is still out on whether expensive, genetically modified (GM) commercial crops will be cost-effective for farmers in the long run, according to a study conducted for the Biotechnology Industry Organization (BIO) and released in Washington last week.

Genetically modifying corn to produce the *Bacillus thuringiensis* (Bt) toxin, which kills the destructive European corn borer, has increased crop yields. US farmers planting such Bt corn gained \$72 million in 1997 compared with conventional varieties.

But the following year, when three times more acreage was planted with Bt corn, growers lost \$26 million as pest infestation levels were low and the price of corn dropped to well below average levels.

Because these factors vary from year to year, the economics of GM crops will only become clear in the long term. Based on historic trends in pest infestation, farmers planting Bt corn can expect three non-paying years every decade. The study, based on the first two to three years of commercial field usage, was conducted

for BIO by the National Center for Food and Agricultural Policy, a non-profit-making institute based in Washington.

Reductions in pesticide use were slight, too: although 18 per cent of US corn planted last year was of the Bt variety, the use of insecticide dropped on only 2.5 per cent of corn acreage. But Bt cotton farmers fared better, gaining \$92 million in net income in 1998 by planting GM crops. Potato farmers, meanwhile, have yet to adapt to large-scale use of Bt strains as they still need to apply insecticide to control other pests.

T.R.

Germany aims for younger professors

[MUNICH] The German University Rectors' Conference (HRK) is proposing to create short-term 'qualification professorships', in an attempt to lower the age at which researchers can become tenured professors, and to offer an alternative to the much criticized *Habilitation* system.

Possible changes in the federal law are to be discussed by a committee of academic experts set up by research minister Edelgard Bulmahn. The committee will also suggest how academic employment rules could be relaxed, for example, to give universities more freedom to pay professors on performance.

Although *Habilitation* — or 'equivalent experience' — is required by law for a professorship, many academics consider it an archaic anomaly. US-style assistant professorships do not exist in Germany. *Habilitation* chains young PhDs to established professors, whom they must serve as 'scientific assistants' to get teaching and research experience. A *Habilitation* must also submit a thesis, known as an *opus magnum*.

The procedure, which exists only in German-speaking countries, is criticized for not giving young scientists independent research experience, and for taking too long: the average age at qualification is 40.

Equivalent experience — such as research, publications and teaching — is now recognized by many science departments, but is difficult to obtain in Germany, where few positions are available for independent research.

Qualification professorships would offer an alternative career path more in tune with international norms, says the report of the HRK's research committee, approved last week by the HRK's plenary council. Young scientists should become eligible to take up a university chair not later than ten years after graduating, it says.

The report says that the positions should be advertised openly, to attract those doing PhDs or postdocs abroad. Scientific assistantships leading to *Habilitation* are too often appointed in-house, contributing to a lack of transparency in promotions and a lack of mobility within the scientific community, says the report.

The qualification professorships should be given to those completing their PhDs at a different university. They should be four-year appointments, with the possibility of a two-year extension.

Klaus Landfried, president of the HRK, says that the positions would be similar to US assistant professorships, and universities should consider them as part of a tenure-track system. "Position holders would be responsible for their own teaching and research programmes, and applying for their own research grants," he says.

The idea builds on the new Emmy Noether programme of the research council, the Deutsche Forschungsgemeinschaft (DFG), which supports young scientists for two years of research abroad and a further three years in Germany. These awards are only available to scientists under 30 years of age (see *Nature* 399, 186; 1999). The HRK and the DFG are "of the same mind" on the need for a career structure for young scientists, says Landfried.

Most scientists have welcomed the proposal, but a few have reservations. Michael Schreiber, professor of theoretical physics at Chemnitz University and a member of the Hochschulverband, the university professors' association, says that *Habilitation* should be maintained in parallel to the proposed new system because it is the best proof of competence for a professorship.

He also points out that scientists can gain their *Habilitation* in their mid-thirties, "if they are clever enough to select a research department where *Habilitation* goes through quickly," such as his own. The HRK's report says that *Habilitation* should remain an option for those departments with faith in it. But most believe that, given alternative ways of proving competence, *Habilitation* will slowly fade away.

Bulmahn's committee will present its report on the general relaxation of various legal restrictions on academic employment in

April next year. One member, Hans Meyer, a lawyer who is president of Humboldt University, Berlin, says that there is little resistance to the idea of 'qualification professorships', but more resistance to other ideas, particularly performance-related pay for professors.

The Hochschulverband, for example, fears that decoupling academic appointments from their *Beamter* (civil servant) status, which would allow universities to vary the pay of professors, would deprive them of the academic freedom that the status guarantees.

Despite this, Meyer predicts that reforms allowing qualification professorships and performance-related pay will be in place by 2002 or 2003, the end of the coalition government's term. Most parliamentarians and state governments are in favour.

But Wolfgang Herrmann, president of the Technical University of Munich, says that universities should not delay action to help young scientists begin independent careers. His university is preparing to convert scientific assistant posts to independent assistant professorships, following amendments to the Bavarian state law on universities this year that give universities more control over their spending. But only a change in the federal framework law on universities would allow these assistant professors to teach their own courses, rather than the courses of an established professor.

Alison Abbott

UK physics postdocs turn to jobs in industry

[PARIS] Many British physics postdocs are waiting for a permanent faculty position that will never come, concludes a report released last week by the Institute of Physics. The report says that increasing numbers of physicists are heading into industry, notably the booming high-tech sector, banking and computing companies.

The report, *Career Paths of Physics Postdoctoral Research Staff*, was based on 684 questionnaires from postdocs graduating between 1988 and 1995. It finds that only one in five contract researchers gain permanent university employment, although 61 per cent identify this as one of their goals.

The survey finds that about a third of the respondents are employed in the private sector. Given the choice, half of these now say that they would not repeat their postdoc experience. But, while many postdocs lament not finding a permanent university position, few need to worry about unemployment — their unemployment rate is 1.3 per cent.

Philip Diamond, manager of higher education and research at the institute, says the report makes it clear that physicists in PhD and postdoc positions need more career counselling. "If they're not going to get an

To think of what I gave up as a Physics postdoc: the fruitless wait for the faculty position...the queue for the photocopier...the permanently broken coffee machine...



academic post they need to know what is out there," says Diamond. Only one in four postdocs could report any examples of career development in their university.

"The shift in employment indicates an increased willingness on the part of new PhDs to pursue non-academic employment," says an earlier report from the American Institute of Physics. "The strong economy, coupled with heightened technical sophistication, has [also] increased the number of employers who are seeking the skills a physics doctorate has to offer."

Heather McCabe

Russian environmental researcher falls foul of security services

[MOSCOW] A prominent Russian ecologist has had his passport seized and his laboratory sealed by the country's security services. Confidential reports and classified maps were removed from the Vladivostok apartment and laboratory of Vladimir Soyfer, who has been working on nuclear pollution in the far east of Russia for the last 40 years.

The warrant for the search states: "Since Soyfer's actions pose a threat to the state and military security of the Russian Federation, it seemed necessary to analyse his correspondence." Material confiscated is being studied by military counter-espionage officials. This may lead to Soyfer being charged with revealing state secrets.

Security official Aleksander Kazakov has denied that he knew Soyfer's apartment was being searched, although a check on Soyfer's laboratory had been planned as a routine procedure. The Federal Security Service (FSB) have not confirmed the reports.

Kazakov says that Soyfer was informed of the time of this check, but "he did not wish to be present". But Soyfer denies this, pointing out that he could not be present as he was being treated for diabetes at the Moscow hospital of the Russian Academy of Sciences.

According to the investigation, Soyfer kept secret documents in his laboratory safe that should have been given back to officials, as well as illegal photocopies. The warrant accuses him of negligence under the Russian Criminal Code, although this does not carry a prison sentence.

A letter distributed by Aleksei Yablokov, chairman of the Social-Ecological Union, suggests that the secret services may be trying to prove that Soyfer has passed secret information to foreign agents.

Counter-admiral Nikolai Sotskov, chief of the Pacific Navy's counter-espionage department, denies this. "Neither our navy, nor the FSB, prevent anyone from monitoring the environment," he says.

But Soyfer has had run-ins with the navy before. The author of over 200 scientific articles on radioactive pollution in the region, he has been studying the consequences of a nuclear submarine accident in the bay of Chazhma in 1985, and of radioactive waste dumped in the Japan sea. This led to a confrontation with Pacific Navy commander admiral Mikhail Zakharenko.

"Ecological investigations attract the attention of the special services, because by analysing the structure of pollution it is easy to tell what kind of secret objects — especially nuclear facilities — are nearby," says Nikolay Ponomarev-Stepnoy, vice president of the Kurchatov Institute.

Carl Levitin

Merger of top Californian medical schools turns sour

[SAN FRANCISCO] The merger of hospitals of the University of California at San Francisco (UCSF) and Stanford University, barely two years old, may be dissolved after the disclosure of projected, unexpected losses of at least \$60 million in the current fiscal year.

The financial situation, which is being blamed on management weaknesses, is so dire that UCSF's faculty leaders last week asked the UC Board of Regents to cancel the hospital partnership.

The merger was designed to meld the clinical operations of the two universities into an academic powerhouse. But it has become a financial nightmare that some officials say is threatening research and teaching.

While faculty members attempt to meet its mission, physicians say they are under such pressure to produce revenue that research is being curtailed, if not eliminated.

"It is time for the regents to exercise their fiduciary responsibility to save the academic mission before it is gutted to save a dysfunctional corporate dream," says Warren Gold, a physician who chairs the UCSF Faculty Association, urging the dissolution of the merger. "The longer you wait before acting, the more difficult it will be to get UCSF back on track."

The regents have decided to support the merger because its goals are too important to let it fail. But the hospital marriage was dealt another blow last week by a legal battle between state authorities and hospital officials. Officials of the hospital corporation, arguing that they were not under state jurisdiction, refused to let the state auditor examine financial records. The auditor then prepared to subpoena the hospital corporation's financial records, while threatening criminal charges against corporate officers who blocked access.

Since the merger, in late 1997, the hospitals — UCSF's facilities in San Francisco and Stanford's in Palo Alto — have been governed by an independent, nonprofit corporation, with a board of UCSF and Stanford representatives.

The auditor and state legislators representing the San Francisco area are concerned about the hospital corporation's growing deficits, including this year's projected \$60 million loss and estimated losses next year of \$135 million.

University officials attributed the losses to mismanagement of the merger and reductions in government support. Officials at Stanford, a private institution, have been watching the conflict nervously, as the university entered the merger under the agree-



Stanford University: officials are reluctant to open medical-centre accounts to state audit.

ment that the hospital corporation would not disclose financial records.

Last Friday (16 July), a compromise was worked out at a closed-door meeting that would open the hospital corporation's financial records to the state auditor. Had the compromise not been reached, the merger could have been cancelled within weeks.

But the hospital marriage remains under threat, as critics doubt that Stanford will allow a full audit that will satisfy the state. As the compromise was being reached, Eugene D. Bauer, Stanford's vice-president for medical affairs and medical school dean, said that opening the financial books "was a violation of the spirit" of the merger.

At UCSF, the impact of the current situation is most severe for physicians who also work in research — faculty considered crucial to the university's reputation. "This is a tremendous worry," said Svein Oie, a pharmacology researcher who is chairman of the UCSF Medical School's Academic Senate.

"The real problem is with the young people; they find it harder and harder to do research," says Oie. "Unless we can reduce the clinical-care burden on the MD/PhD faculty, I don't know if we can resolve the problem."

"For those of us trying to do lab work, it is more and more difficult," said an MD/PhD in the UCSF anaesthesia department, asking not to be named. "Working in the lab is becoming secondary, which will make it harder for academic promotion."

At Stanford, the crisis has prompted some faculty hired for scholarly work to leave, with their positions being filled by clinicians.

"We are in quite a bit of trouble," admits Kenneth L. Melmon, chairman of Stanford Medical School's Academic Senate, a physician and associate dean for postgraduate medical education. "The constitution of the faculty is changing."

At UCSF, the crisis is pitting physicians against basic science researchers. Some basic scientists have even argued that clinical programmes aren't needed — something the physicians contest sharply.

Rex Dalton

Britain faces up to transparency review

[LONDON] British academics are reacting warily to the prospect of increased paperwork and closer scrutiny of their activities as a result of the first wave of a 'transparency and accountability review' of research.

The review is being introduced into eight pilot universities following the government's comprehensive spending review last year. This made more money available to universities on condition that they accept increased public scrutiny of how the money is spent.

The Treasury-driven review requires universities to cost the activities of staff by department, subject and type of research sponsor. The government wants greater transparency of university research, on which it spends £1.7 billion a year.

"Improving transparency is an important step in maintaining the health of university research," said John Taylor, director general of the research councils, last week. He was commenting on a report outlining how universities could implement a uniform approach to costing research.

Taylor promised that the methodology proposed would not impose "requirements that are unnecessarily burdensome". But the academic community is concerned it will mean more paperwork and closer scrutiny of spending decisions.

"Of course they are suspicious," says Robin Jackson, policy adviser for the Com-

mittee of Vice-Chancellors and Principals. "This is an additional requirement on an overburdened and under-funded system." But he points out that it is "not an exercise to determine a shortfall in funding".

Heather Williams of the Higher Education Funding Council for England, and also a member of the steering group that commissioned the report, says the methodology "is simple to implement and gives reasonably robust, auditable figures without an administrative burden".

Some argue that the review cannot determine whether the full costs of research are being met by research funders, as they would like, as it fails to require reporting of the number of hours worked by academics.

"This is where people magic money out of thin air when they say the research base is well funded," says Peter Cotgreave, director of the lobby group Save British Science. "When the issue of full overhead costs for grants comes up, people say that, because the research is being done, there must be enough [money]. The truth is that people work longer hours to get it done."

There is concern that the transparency review could work against universities if it is found that teaching grants are being spent on research. But this seems unlikely, given the pressures on teaching resources within higher education. The recent expansion of stu-



The novelist, Terry Pratchett (centre) has received an honorary degree at the University of Warwick for the science content of his 'Discworld' books. Ian Stewart (left) and Jack Cohen (right), his co-authors on *The Science of Discworld* (Ebury), were made 'honorary wizards of the Unseen University'.

dent numbers has already seen a decline in the resource allocation per student.

There will be a three-stage implementation of the proposals. New costing standards should be reported by January 2001 for eight pilot universities, and by summer 2001 for a further 30 universities.

Natasha Loder

White House cool on obtaining human embryonic stem cells

[WASHINGTON] The Clinton administration last week distanced itself from an impending recommendation by a presidential advisory commission that the government should fund both the derivation and the research use of human embryonic stem cells.

The recommendation is expected from the National Bioethics Advisory Commission (NBAC), which met in Cambridge, Massachusetts, last week to finalize a report to President Bill Clinton on the ethics of stem-cell research. The report may foment a Congressional debate that pits opponents of abortion against advocates of stem-cell research (see *Nature* 399, 292; 1999).

But the White House seems to be trying to avoid an open conflict. Advocates of stem-cell research hope it will yield cell and tissue therapies for diseases such as Parkinson's and juvenile diabetes. Opponents decry the research because the derivation of stem cells requires the destruction of embryos.

In a statement issued after NBAC met, the White House press secretary said the administration "recognizes that human stem-cell technology's potential medical benefits are compelling and worthy of pursuit, so long as the research is conducted

according to the highest ethical standards".

But the statement added an implicit repudiation of the NBAC's recommendation that the government should finance the harvesting of stem cells from embryos left over after fertility treatments. It said that no legal action is necessary now "because it appears that human embryonic stem cells will be available from the private sector".

If the administration were to embrace a recommendation that it fund the collection of stem cells, it would have to lobby Congress to overturn or liberalize a law that bans government funding of research in which embryos are destroyed or discarded.

The government position is that federal funds may finance research on stem cells but not their harvesting, which must be done with private funds. This was delivered in a decision by the legal counsel of the Department of Health and Human Services in January (see *Nature* 397, 185; 1999).

But the authors of the ban on embryo research contest this interpretation. They may try to rewrite the ban to apply explicitly to stem-cell research when the appropriations bill to which it is attached is finalized this summer. If they do so, it is not

clear which side would prevail in Congress.

Frank Young, a commissioner of the Food and Drug Administration under Ronald Reagan's presidency, said: "To say, on the one hand, that you cannot support the deliberate destruction of living human embryos to harvest their stem cells, but that you will, on the other hand, pour millions of taxpayer dollars in support of research that you know can only take place using materials derived from that destruction, is an exercise in sophistry, not ethics." Young was speaking on behalf of a group of abortion opponents called Do No Harm: The Coalition of Americans for Research Ethics.

But Joseph Cerquone, a spokesman for the Patients' Coalition for Urgent Research, said his organization was "heartened" by the NBAC report and the White House reaction.

A spokeswoman for Harold Varmus, director of the National Institutes of Health, said last week that guidelines under which investigators funded by the agency may proceed with stem-cell work will soon be published. "We're being careful with them," she said, adding that they are already being discussed with the Department of Health and Human Services.

Meredith Wadman

Plan to create agency for US weapons labs under strong criticism

[WASHINGTON] Leading congressmen from both parties in the US House of Representatives are fighting a plan to reform the nuclear weapons laboratories by placing them under a semi-autonomous agency within the Department of Energy.

The plan is championed by Senator Pete Domenici (Republican, New Mexico). It has wide support in the Senate and the partial backing of energy secretary Bill Richardson. But opponents say it is insufficiently radical to address security concerns in the wake of allegations of Chinese espionage at the Los Alamos laboratory in New Mexico, and would in fact protect the managers responsible for these concerns.

The House subcommittee that controls the energy department's budget proposed last week that it should be cut severely next year — and that an additional \$1 billion for the weapons laboratories should be held in reserve until the management structure of the labs is reformed. Ron Packard (Republican, California), chair of the subcommittee, is believed to want to use the measure to negotiate more radical reforms than those proposed by Domenici.

Two days earlier, John Dingell (Democrat, Michigan), the senior minority member of the House Commerce Committee, had led a separate bipartisan attack on Domenici's plan at hearings of two House subcommittees. Dingell described the proposal for a semi-autonomous agency as "using gasoline to extinguish a fire".

Dingell and other Democrats compared Domenici's proposal to the Atomic Energy Commission, which operated the weapons laboratories until the mid-1970s. Although remembered fondly by some physicists, the commission is regarded by Dingell as an example of closed government. "It was one of the most arrogant and incompetent government agencies I ever saw," he said.

Republicans on the panel, meanwhile, attacked the management abilities of the energy department. They said they would prepare legislation to restructure it.

Packard's subcommittee proposed large across-the-board cuts in the energy department budget, including cuts of \$525 million in President Bill Clinton's request for the nuclear weapons programme, \$250 million in energy supply programmes and \$350 million in environmental clean-up.

The House is now preparing its own legislation to reform the department. But observers say that Domenici, as a senior member of several committees that oversee the labs, is in a strong position to defend them from radical changes. **Colin Macilwain**

\$3m deal launches major hunt for drug leads in Brazil

[SÃO PAULO] The pharmaceutical multinational Glaxo Wellcome and a small Brazilian biotechnology company signed a \$3.2-million contract last week to screen up to 30,000 compounds of plant, fungus and bacterial origin from several regions in the country. The initiative is described as the world's largest natural product sampling and screening programme.

As part of the three-year deal, the companies have agreed that one-quarter of any royalties arising from successfully exploited patents will be used to support community-based conservation, health and education projects. This reflects recent heated debates over how compensation should be paid by companies that turn 'indigenous knowledge' into commercial products.

According to Jorge Raimondo, Glaxo's regional director for Latin America, a further 25 per cent of the royalties will go to the university group responsible for isolating and identifying the product in question. Glaxo Wellcome will pay for all research and development costs in Brazil. The research will focus on compounds found in the Amazonian and Atlantic rainforests, and Glaxo Wellcome will have an option to license any product arising from it.

Raimondo says the agreement reflects the company's view that, despite the increasing use of combinatorial chemistry to screen large quantities of artificially synthesized molecules, there is still much to be learned from traditional remedies. "One of our main interests, for example, is the way that some of these compounds control pain."

Officials from Glaxo and the Brazilian company, Extracta, point out that the agree-

ment has been made possible by — and conforms with — new legislation on intellectual property introduced in Brazil after a fierce political debate on the protection of genetic material (see *Nature* **392**, 538; 1998). They say that it also conforms with the United Nations Convention on Biological Diversity, which specifies that the interests of local communities must be taken into account.

"With both Brazil's new patent law and the country's rich biodiversity, the moment is now appropriate to invest in Brazilian science," says Padraic Ward, Glaxo's manager for business development in Latin America.

Antonio Paes de Carvalho, Extracta's director-general and a biophysics professor at the Federal University of Rio de Janeiro, says Brazilian researchers often complain that multinationals do not invest in local research. "This contract opens the way for similar companies to join other university centres," he says. "It is the beginning of a change in attitude."

Paes de Carvalho says the Glaxo-Extracta contract is different from the \$1-million screening deal signed a few years ago by Merck and Costa Rica's INBIO. In contrast to that agreement, scientists in Brazil will work on the identification of molecules and the study of their biological properties, rather than just processing plant extracts for analysis abroad.

According to Raimondo, Glaxo is particularly interested in molecules with potential antibiotic, anti-inflammatory and possibly antifungicide properties. Paes de Carvalho says that one of his goals is to create a database of all the samples and molecules screened. **Ricardo Bonalume Neto & David Dickson**

Outspoken science minister ousted in reshuffle

[SÃO PAULO] Brazilian president Fernando Henrique Cardoso last week replaced the science minister, Luiz Carlos Bresser Pereira, in a ministerial reshuffle. The job went to diplomat Ronald Sardenberg, a personal friend of the president, and previously secretary for special projects and secretary for strategic affairs.

Bresser Pereira, an economist, had been science minister for the first six months of Cardoso's second term, having replaced José Israel Vargas, who held the

post for a record six years. Bresser Pereira was known among scientists for a certain impetuosity and for making controversial statements. He aroused considerable animosity among researchers in northeast Brazil with comments about the quality of the science done there.

Whether or not the minister's style played a part in his replacement, the reshuffle was intended primarily to redistribute favours among the political parties that support Cardoso. For a while the science

ministry was expected to go to a member of one of these parties, but Cardoso eventually decided to appoint someone whom the scientific community could be comfortable with.

Cardoso has said that he considers the ministry's role to be "strategic". "We'll see what happens when the budget for 2000 appears," comments the new president of the Brazilian Society for the Advancement of Science, biochemist Glaci Zancan, from the Federal University of Paraná.

R. B. N.

Brookhaven's heavy ion collider goes live

[LONG ISLAND] Physicists at the Brookhaven National Laboratory on Long Island last week sent gold ions around the first ring of the Relativistic Heavy Ion Collider (RHIC), a \$600 million nuclear-physics facility built to create and study quark–gluon plasma.

Next, the RHIC team will send ions in the opposite direction around a second, 3.8-km circumference, ring, opening the way for the facility's experimental programme to start this November.

The programme is the culmination of a two-decade quest by US nuclear physicists for an experiment powerful enough to explore the behaviour of quarks and gluons when nuclei collide at energies high enough to wrest these particles from their usual status as the constituents of protons and neutrons.

The tunnels that carry RHIC's rings, as well as some of the equipment which will feed ions into them, were constructed 20 years ago as part of a particle physics experiment, called Isabelle, to study proton collisions. But in 1983, high-energy physicists abandoned Isabelle for the more ambitious Superconducting Super Collider — itself abandoned by Congress ten years later.

The demise of Isabelle led Nick Samios, who had become director of Brookhaven in 1982, to work with nuclear physicists to create a heavy-ion collider. Samios, who was ousted as director of Brookhaven two years ago, can now watch his dream reach fruition as an ordinary physicist working at the laboratory.

RHIC will serve as the main focal point for the Brookhaven laboratory, which has been in turmoil following a leak of radioactive tritium from the fuel storage tank of its research reactor. The reactor has been closed since the leak was discovered in 1997 (see box).

Almost 1,000 investigators will use the four detectors at RHIC — two large and two small — to study the character of the quark–gluon plasma. This plasma is expected to be created when gold ions collide at combined energies of 40 TeV, or 200 MeV for each proton or neutron involved in the collision.

The detectors will use differing methods to detect the byproducts of the collision, reflecting uncertainty about what clues will betray the existence or the nature of the plasma. "We're talking about the interaction of hundreds of quarks and gluons," says Samios. "There's lots of speculation about what will happen. The problem is that you don't know what you are looking for."

Broadly speaking, one of the large detectors, called Phenix, will identify rare by-products of the plasma explosion, such as the J/ψ meson, while the other, Star, finds the tracks of the more common ones. Phenix has cost about \$90 million and Star \$70 million; each involves about 400 investigators.

Two more specialized detectors, Brahms



Collision course: rings have been placed inside a tunnel intended for ill-fated Isabelle accelerator.

and Phobos, each involve 50 investigators and cost about \$8 million.

Japan and Russia have been heavily involved in the design and construction of the RHIC detectors — their combined contribution is worth about \$40 million, according to project director Satoshi Ozaki. But there has been little European involvement. European

physicists are planning their own experiment, called Alice, using the Large Hadron Collider when it opens in six years' time at the European Laboratory for Particle Physics (CERN).

The 200 MeV energy for each nucleon in a collision at RHIC is ten times the energy of the most powerful such experiment so far conducted at CERN, and four times the minimum energy that should smash the colliding nuclei into a plasma of quarks and gluons. Physicists are confident that plasma will be created at RHIC, but unsure of the characteristics of the transition to plasma. "The phase transition is a subtle thing to observe," says Tom Ludlam, associate director of the project.

The rewards of finding the plasma, however, are great. RHIC could recreate the conditions thought to characterize the Universe for the first fraction of a second of its existence, before quarks and gluons coalesced into protons and neutrons. The discovery of the plasma would confirm an important prediction of quantum chromodynamics, and its characterization could give a new insight into the nature of matter.

Colin Macilwain

Leading a laboratory out of the mire

[LONG ISLAND] The man trying to steer Brookhaven National Laboratory out of the tempest that consumed it when the local community reacted furiously to a radioactive leak from the laboratory is John Marburger, former president of the nearby State University of New York at Stony Brook (*Nature* **386**, 3, 1997).

Marburger is a soft-spoken administrator whose style contrasts to that of Nick Samios, the brilliant, pony-tailed physicist who led Brookhaven for 15 years. Samios, and the university consortium which managed the lab for the Department of Energy, departed soon after the discovery of the leak.

In management terms, "this laboratory experienced a catastrophe in 1997," says Marburger. "The quality of science in the lab was so good that the Department of Energy was willing to accept an obsolescent management style. People were oblivious to the actual state of the management systems here."

With much of the day-to-day management of science

delegated to Peter Paul, the deputy director, Marburger is concentrating on the lab's external relations, which he says are improving.

But a decision to restart the High Flux Beam Reactor — one of three important user facilities at Brookhaven, closed since the discovery of the leak from a basin storing its spent fuel — is not expected until early next year, or later. "It's a year later than I expected, and I'm disappointed in the schedule," says Marburger.

Given the opposition of local congressman, Michael Forbes (who this week switched from Republican to Democrat), and tight budgets, Marburger finds it hard to muster much optimism that the reactor will reopen at all, though "it is hard to imagine it not reopening, because it is so badly needed".

Since Brookhaven Science Associates (BSA), a new contractor involving the research consultancy Battelle, and Stony Brook, took over operation of Brookhaven last year,

rumours have circulated that Battelle will cut staff and steer the lab towards technology which it can sell.

Marburger denies that either is being considered. "Battelle is delighted to be associated with the basic science at this lab," he says, adding that a shift to technology "would be a disaster. As president of BSA, I intend to maintain the fundamental characteristics of this laboratory."

As for staff, Marburger says the lab needs to expand its 3,100 staff, to generate the research income needed to cover its overheads. "I'd like to increase the staff to 4,000," he says.

"We believe our basic science capability is not being fully exploited," Marburger says. "Our basic strengths are in fields that are a little old-fashioned." He declines to identify which fields, but says that growth areas should include structural biology, imaging technology, chemistry and materials science.

C.M.

German call to limit use of genetic data by insurance companies

[MUNICH] Germany's science funding agency, the Deutsche Forschungsgemeinschaft (DFG), has called on employers and the insurance industry to voluntarily minimize the use of genetic data for employment and insurance purposes.

In a report, *Opportunities, Limitations and Consequences of Human Genome Research and Predictive Genetic Diagnostics*, published last week, the DFG recommends that genetic tests be limited to medical purposes. But it concludes that there is as yet no need for new legislation, as both the number of diseases that can be predicted genetically and the number of patients involved is small.

Genetic tests should not be made a condition for taking out an insurance policy, according to the report, although prospective policyholders are obliged to state the risk of genetic diseases predicted by previous tests.

The DFG's recommendations are consistent with the insurance industry's self-imposed standards, according to Christian Weber, a spokesman for the German association of private health insurers.

Neither new legislation nor a moratorium

on the use of genetic information is needed in Germany, Weber argues.

Congress bid to extend research tax credit

[WASHINGTON] A key committee in the US House of Representatives last week extended the research and development tax credit for five years. If the provision survives — it is contained in an \$864 billion House tax-cut package and a \$792 billion Senate companion — it would mark the longest extension of the credit in its 18-year history.

The credit, which costs about \$2.5 billion but is projected to rise by \$200 million yearly to 2001, gives companies a 20 per cent tax credit on new research expenditures. But although House and Senate Republicans have backed the five-year extension of the credit, its survival is uncertain. President Bill Clinton opposes the large tax cuts congressional Republicans are proposing, and may veto any bill that contains them.

Drug could slash rate of perinatal HIV

[WASHINGTON] A new drug regimen has been found to cut perinatal transmission of HIV by 47 per cent compared to the standard regimen in developing countries. A joint

US-Uganda study found that if Nevirapine, a non-nucleoside reverse-transcriptase inhibitor developed by Boehringer Ingelheim Pharmaceuticals, is used in place of AZT, mother to child HIV transmission fell to 13 per cent. The rate with short-course AZT is 25 per cent.

Nevirapine treatment also costs much less. A mother takes a pill once during labour, and the baby is fed the drug as a syrup once in the three days after birth. This costs US\$4 for the two doses, as against \$268 for short-course AZT. Officials predict that use of Nevirapine could prevent 300,000 to 400,000 newborns annually from contracting HIV.

Jobs face axe at British arable crops institute

[LONDON] Britain's Biotechnology and Biological Sciences Research Council announced last week that it is withdrawing funding from the Long Ashton site of the Institute of Arable Crops Research (IACR), owned by the University of Bristol. The institute's research activities will be focused on its main centre at Rothamsted in Hertfordshire. Some of the Long Ashton work will be transferred to Rothamsted, which will be given a £19 million (US\$30 million) refurbishment over the next four years.

The labour union representing staff at Long Ashton, the Institute of Professionals, Managers and Specialists, has vowed to fight the decision. It says that 160 members of the research staff at the laboratory, which employs 230 staff, will lose their jobs. But Ian Crute, the director of the institute, argues that the restructuring is "science-led" and represents "a major opportunity for IACR to develop a unique integrated research resource for UK agriculture".

Japan sets five-year genome research plan

[TOKYO] Japan announced last week a five-year plan to increase funding to speed up research into the human genome and to boost the biotechnology industry. It has already announced plans to create a database of single nucleotide polymorphisms believed to be unique to Japanese people (see *Nature* 400, 6; 1999).

The government hopes to create pharmaceuticals and diagnostic techniques by concentrating efforts on the analysis of around a third of the human genome in the next few years. The strategy was agreed by the Science and Technology Agency, the Ministry of International Trade and Industry, and the Ministry of Education, Science, Sports and Culture.

Court bid to throw out animal-rights lawsuit

[WASHINGTON] Lawyers for the US government last week urged a federal court to dismiss a lawsuit that would compel the government to extend the Animal Welfare Act to cover some 23 million research rats, mice and birds (see *Nature* 400, 197; 1999).

In a US District Court, lawyers for the United States Department of Agriculture (USDA), which enforces the 1966 law, argued that the animal-rights activists who filed the lawsuit in March do not have legal standing to sue. The government also said that the court should stay the lawsuit as premature, as the plaintiffs have petitioned USDA to include the rodents in the law's protections.

US biomedical labs to reach out to schools

[WASHINGTON] The Howard Hughes Medical Institute has awarded \$12.7 million to 35 US biomedical research institutions to help them strengthen links to local high schools.

The four-year grants are intended to help the research institutions to share their laboratories and research skills with the local community, and to attract a broad range of students into medical careers. For historical reasons, many of the top medical schools in

the United States are located in socially-deprived areas of inner cities.

US loses its 'foremost advocate for science'



[WASHINGTON] George Brown (Democrat, California), the senior minority member of the Science Committee of the House of Representatives and probably the most respected advocate for science in either chamber of the US Congress, died

on 15 July at the Bethesda Naval Hospital, Maryland. Brown (above), aged 79, had contracted an infection after undergoing heart valve replacement surgery in May.

Brown had served 18 two-year terms in the House. Although he had a degree in industrial physics, he described himself as a layman and said he pursued science and technology issues as a means to peace and social justice.

"America has lost its foremost science advocate, a statesman and a tremendous human being," said Jim Sensenbrenner (Republican, Wisconsin), chairman of the Science Committee. Brown will be succeeded as senior minority member of the committee by Ralph Hall (Democrat, Texas).

Unfamiliar citations breed mistakes

Sir—Methods for assessing the quality of research are increasingly based on impact factors and citation analysis. But there is concern about the accuracy of science citation^{1–7}, and suspicion has been cast on the impartiality of citation analysis^{2–4}.

As well as conscious bias^{5,8}, it has been suggested that there is a tendency to make more citation errors in names from unfamiliar languages⁴. We have examined the tendency to make errors when typing strings of familiar and unfamiliar names. These experiments, together with an analysis of the numbers of authors per paper in different journals, show why there is a positive association between journal quality (impact factor) and rate of apparent undercitation⁴.

To see whether language familiarity and the length of name strings influenced typing accuracy, we asked 36 undergraduate zoology students to type name strings that consisted of one, three or six Finnish (unfamiliar) or English (familiar) names that were projected on a screen for 8, 24 or 48 seconds, respectively. All subjects were native English speakers with no knowledge of Finnish. The error rate in unfamiliar names increased significantly faster with the length of the name string than the error rate in familiar names (Fig. 1).

We then calculated the mean number of authors per article for each of 65 journals, and plotted this number against the journal's impact factor. We found a strong positive correlation (Fig. 2). Our experimental results (Fig. 1) show that typing errors occur at a greater rate in long strings of non-English names. Therefore, the positive correlation between journal impact factor and number of authors

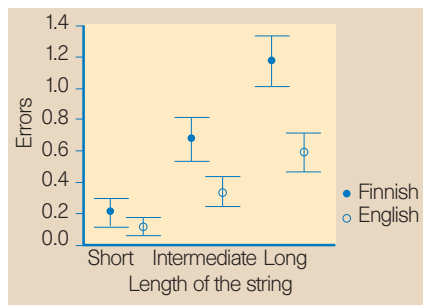


Figure 1 Number of errors (mean $\pm$ 1 s.e.m.) in short, intermediate and long name strings for Finnish and English names. Errors are ln-transformed. There is a significant interaction between language and length of the name string on error rate (repeated measures ANOVA $F_{1,50,52,33} = 6.44$, $P = 0.007$). (To correct for a violation of the sphericity assumption, the degrees of freedom are Greenhouse–Geisser adjusted.) Short name string was one name totalling 5 to 8 characters; intermediate string, three names totalling 15 to 24 characters; and long string, six names totalling 30 to 48 characters. There was no difference in the total number of characters between Finnish and English name strings (ANOVA $F_{1,54} = 0.00$, $P = 0.953$). Finnish names with non-English characters (Å and Ö) were not used. All subjects received all combinations 20 times and each name was used only once.

(Fig. 2) will inflate the effect of mis-citation, and contribute to the positive association between journal impact factor and rate of apparent undercitation⁴.

Our results confirm suspicions that citation analysis is biased. Mis-citation of unfamiliar names is serious, because almost half of all scientific publications come from English-speaking countries¹, so non-

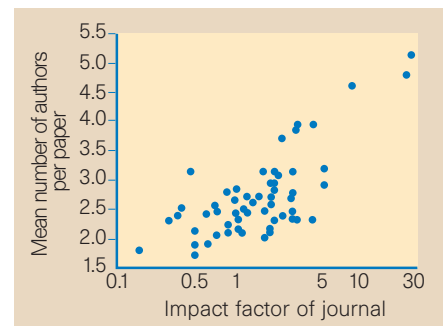


Figure 2 Correlation between a journal's impact factor and the mean number of authors per paper. Pearson's $r = 0.706$, $n = 65$, $P < 0.001$. Journals are a random sample of biological/ecological/evolutionary journals from the Biological Sciences Library of the University of Western Australia. The first 30 papers in 1998 in each journal were analysed. When there were fewer than 30 papers during that year, additional papers from 1997 were included. Impact factors are from the 1997 Science Citation Index.

English-speaking countries will suffer more mis-citations. Our results suggest that undercitation^{3,4} could be an artefact of mis-citation rather than true undercitation, and so cast doubt on the objectivity of citation analysis.

Janne S. Kotiaho, Joseph L. Tomkins, Leigh W. Simmons

Evolutionary Biology Research Group, Department of Zoology, University of Western Australia, Nedlands, Western Australia 6907, Australia

1. May, R. M. *Science* **275**, 793–796 (1997).
2. Tregenza, T. *Nature* **385**, 480 (1997).
3. Paris, G. *et al.* *Nature* **396**, 210 (1998).
4. Kotiaho, J. S. *Nature* **398**, 19 (1999).
5. Barreto, G. R. *Science* **276**, 882–883 (1997).
6. Price, N. C. *Nature* **395**, 538 (1998).
7. Brown, N. L. *Nature* **398**, 555 (1999).
8. Gibbs, W. W. *Sci. Am.* **273**, 76–83 (1995).

Ethics of population genomics research

Sir—A recent News article described a new Swedish genomics company, UmanGenomics, as having a higher ethical nature than another genomics company, deCODE genetics, which was portrayed as having questionable practices (*Nature* **400**, 3; 1999). We do not accept this characterization.

The article described deCODE as a “US-based gene-brokering company” with special rights in Iceland. But deCODE and its 260 employees are solely based in Iceland; the company has already delivered to Iceland the much-needed high-tech jobs that UmanGenomics has promised in Sweden.

From early on, deCODE set high ethical and privacy standards that meet or exceed European Union regulations for human research, under the scrutiny of a national bioethics committee and the Data Protection Commission (DPC) of Iceland. The company has not generated any data from biological material without informed consent granted specifically for each study. All information and biological samples arrive at the company with personal identification numbers encrypted by the DPC. To our knowledge, the current system for protection of privacy at deCODE is unsurpassed (see www.decode.is).

Comprehensive genealogy is an essential tool for population genomics research. It does not yet exist in Sweden but, because Icelanders have been uniquely interested in their genealogy for the past 11 centuries,

deCODE has been able to assemble a computerized genealogy database containing all 270,000 living Icelanders, together with 330,000 of their ancestors.

The News article states that a population-based biobank is to be generated in Iceland under the exclusive control of deCODE, and that Icelandic scientists would have to pay deCODE to carry out basic research. This is incorrect. The Icelandic health-care database would contain medical information with irreversibly encrypted identifiers that, according to database law, can be linked to genealogy and genetic information only with informed consent. The bill grants deCODE the exclusive right to market this anonymous database outside Iceland, but not the exclusive right to use the database. According to the law, the database is to be

controlled by a licensee (deCODE) as well as by oversight and ethics committees and the DPC. Icelandic scientists choosing not to collaborate with deCODE will have the same or better access to patients and their medical records, not less. Those doing non-commercial research will have free access to the database. The law dictates that the data are to be stored so that they cannot be linked to individuals, so the database would not be a biobank as defined by the Swedes.

The article incorrectly gives the impression that UmanGenomics' use of its Medical Bank is going to be with the consent of the community, whereas deCODE's use of a database was opposed by the community. The Icelandic poll actually asked whether people were concerned about the protection of privacy in a centralized medical database, and (surprisingly) only 25% said yes. UmanGenomics was granted permission to use its Medical Bank by government committees and bureaucrats, which is not 'community consent'.

In stark contrast, deCODE obtained its licence to construct and run the health-care database through the democratic process. In 1997, the company suggested the database to the Ministry of Health, which then drafted a bill and placed it on its open website for comment. The government submitted the bill to the parliament (Althingi) in early 1998. Vigorous local debate lasted nine months and included hundreds of newspaper articles, radio and television programmes. Icelanders debated the database bill more than any other bill in the history of the republic. On the eve of the vote on the bill, a poll showed that 75% of the population supported it, and Althingi passed it last December by the same margin.

One premise in the News article is that, because government institutions own 51% of UmanGenomics, the proper use of its Medical Bank is assured. But governments have a bad record on violation of privacy. Further, the health-care authorities and the university that own most of UmanGenomics are mainly concerned with attending to diseases and health. Majority ownership of a genomics company that uses health-care information from clients as raw material may be seen as a serious conflict of interest.

Jeffrey R. Gulcher, Kari Stefansson
deCODE genetics, Lynghals 1, Reykjavik, Iceland 110

Striking the right note

Sir — About 50 years ago, a student was working in the library of the Natural History Museum in London one afternoon, and overheard two rather elderly gentlemen discussing *Nature* as they leafed through a

copy on display. "Do you still read this these days?" asked one. "Not really," replied the other. "I put my copy of *Nature* on the music-stand and play it."

Well he might, had he encountered one of those genome pull-outs you offer us these days, although he might not have understood the coloured musical notation (*Nature* 399, 323–329; 1999).

Alan Longhurst

Place de l'Eglise, 46160 Cajarc, France

Heisenberg, the bomb and the historical record

Sir — Mark Walker's review¹ of my book *Heisenberg and the Nazi Atomic Bomb Project* distorts the record in several ways.

Walker conceals from readers the fact that the book contains documented proof of his own past suppression of crucial evidence. This raises a question of reviewing ethics.

Walker hides the scientific issues relating to Werner Heisenberg's misunderstanding of the critical mass of an atomic bomb which he calculated to be *tons* of uranium-235. Heisenberg's mistake has been fully demonstrated by the publication by Sir Charles Frank in 1993 of the Farm Hall transcripts². Walker has already been reprimanded in an article in *Nature*³ for misleading readers, but he still does not see fit to mention the fundamental Farm Hall evidence in his review. In his own writings, Walker has all along suppressed Frank's explanation to him, in a taped interview of 1985, of the Heisenberg error.

Walker pretends that Heisenberg never seriously considered an exploding reactor-bomb. My book cites German reports analysing such a bomb, including two by a member of Heisenberg's team and a lengthy 1942 synoptic report, as well as Heisenberg's own comments, which show the idea to have been serious.

Any one of these three points should suffice to warn your readers against taking on trust any statement in Walker's review.

Paul Lawrence Rose

Institute for the Arts and Humanistic Studies,
Pennsylvania State University,
University Park, Pennsylvania 16802, USA

Sir — What Walker somehow fails to mention in his dismissive review¹ is that Rose's book includes a devastating critique of Walker's own work on the subject. Re-examining the technical records on which Walker based the central thesis of his 1989 book⁴, Rose shows that Walker misrepresented crucial documents and suppressed essential evidence.

Ten years ago, Walker claimed to have

uncovered archival documents proving that German physicists had "performed the same sort of experiments, had made the same type of calculations, and had come to similar conclusions as the Allies — for example the estimate of explosive critical mass". As Rose demonstrates, this interpretation can be made to seem plausible only by the most convoluted and selective reading of the evidence.

A key 1942 German progress report, for example, cited by Walker to demonstrate that Heisenberg knew the critical mass of a uranium-235 bomb to be small, does contain a parenthetical speculation that with plutonium (then a hypothetical element) a small critical mass might obtain. What is described in the main text of the report, however, is the unworkable multi-ton reactor-bomb first intimated by Heisenberg in 1939. By citing only the parenthetical remark while suppressing the report's substance, Walker transforms contrary evidence into support for his thesis. Inconvenient technical reports by Heisenberg's research assistant Paul Müller elaborating the misguided reactor-bomb concept go unmentioned in Walker's account. An oral history interview with wartime scientific intelligence officer Sir Charles Frank forcefully describing Heisenberg's mistaken estimate of the critical mass is likewise suppressed — Walker cannot have been unaware of it since it was he who conducted the interview. Thus did Walker succeed in "misrepresenting not only the content of the document[s] but also the whole history of the German atomic project".

Reading Walker's review, one would never know any of this, or that such objections had ever been raised before^{3,5,6}. "Rose's book does not really have a conclusion," writes Walker, while suggesting that there is little in it that could possibly interest the readers of *Nature*. Having made the unfortunate decision to review a book that contains discrediting revelations about his own scholarship, Walker might have taken the opportunity to address the disturbing issues it raised. At the least, readers are entitled to a disclosure of self-interest. Instead, Walker chose the dismal principle he has followed before, *si incommoditas est, non est*. Whether a discordant fact, an inconvenient document, or a detailed study putting his claims to the test: if it doesn't fit, it just didn't happen.

Jonothan L. Logan

EPG Research Foundation, 111 East Shore Road,
Manhasset, New York 11030, USA

1. Walker, M. *Nature* 396, 427–428 (1998).

2. Frank, C. *Operation Epsilon* (Institute of Physics, Bristol, 1993).

3. Logan, J. L. & Serber, R. *Nature* 362, 117 (1993).

4. Walker, M. *German National Socialism and the Quest for Nuclear Power: 1939–49* (Cambridge Univ. Press, 1989).

5. Logan, J. *American Scientist* 84, 263–277 (1996).

6. Klotz, I. M. *Nature* 379, 410–412 (1996).

Strengthening biomedicine's roots

The future of biomedicine depends on the strength of its underpinning sciences, including chemistry, physics and mathematics. Therein lie some of biomedicine's most challenging and costly problems, and greatest opportunities.

Ken A. Dill

A key trend of biology and medicine this century has been to reduce the problems of disease to problems of molecular science: to identify the relevant molecules, to purify them, to determine their three-dimensional structures at atomic resolution, and then to create drugs to bind to them, atom by atom. Many of the associated methodological advances in biomedical science are the result of earlier investments in the basic sciences (see Table 1).

Breakthroughs in molecular science have opened a floodgate of new opportunities for curing disease. A decade ago, biologists studied one protein at a time, one gene at a time, and one drug molecule at a time. Now, genome sequences are pouring in at the rate of hundreds of megabases per year. High-throughput screening and combinatorial chemistry are producing 100,000 candidate drug molecules per month. And gene chips can monitor nearly 10,000 genes in a single experiment. We are no longer limited by how quickly we can obtain genome sequences or synthesize molecules. About 20 genomes have already been sequenced, with nearly 100 more in progress, including the human genome. All of them are likely to be solved in less than five years.

But with the acceleration of opportunities comes the acceleration of costs. Obtaining a protein structure has been estimated¹ to cost US\$150,000. So, if there are hundreds or thousands of interesting proteins per genome, we are already on course to spend profitably tens of billions of dollars, using today's proven strategies for curing disease. At this rate, the enterprise of molecular and structural biology alone could easily grow to overwhelm the combined research budgets of the US National Institutes of Health (NIH) (\$15 billion per year) and the US pharmaceutical industry (\$27 billion per year)².

Our challenge now is how to circumvent these costs. Again, molecular science is the key. How can we comprehend molecular structures and functions and create therapies, devices and drugs much faster than we do now? Structural biology is still stuck at the twentieth-century pace: one molecular structure at a time, about 1,000 per year. This means that, from the genome sequences we'll have in the next five years, we'll already have a 100-year backlog of structures we'll want to know.

We need new high-throughput methods to get molecular structures better, faster and cheaper. And we need computational meth-

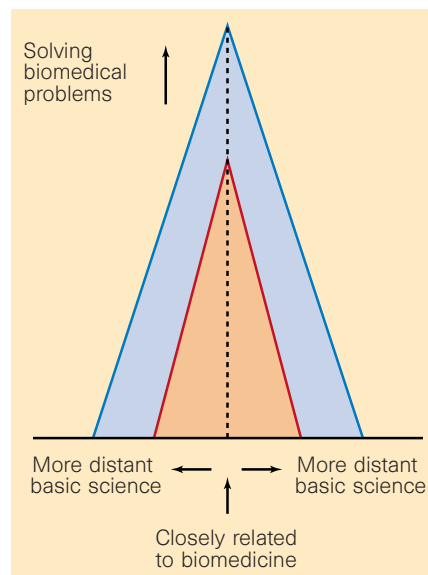


Figure 1 To solve more biomedical problems, we need a broader base in the physical sciences.

ods that could circumvent this bottleneck, to step from genomes to drugs. We need more synchrotron beam lines, phasing methods, more sensitive detectors, and perhaps X-ray lenses for holography. We need to learn to make crystals faster, larger and of higher quality. We need higher-field superconducting magnets for nuclear magnetic resonance (NMR) structure determinations, methods to solve structures of membrane proteins, and new nanotechniques, such as atomic force microscopy, to see molecular assemblies. We need a better understanding of solvation to help design drugs, and of how molecular crowding causes properties of molecules *in vivo* to differ from those *in vitro*. We also need ways to go beyond static structures, to see motions and mechanisms more directly.

This list is only the tip of the iceberg; it focuses on single biomolecules. Similar lists could be compiled for molecular interactions and complexes, genetic networks and cellular functions, and others. And small steps will not do. We need order-of-magnitude advances just to keep up with the opportunities at hand.

It is time to mount a substantial new enterprise to focus on the basic molecular and quantitative sciences that underpin biomedicine. Investing even a small fraction of our resources in the underpinning disciplines would be an investment in better, faster and cheaper ways to cure disease.

The research pyramid

The biomedical research enterprise can be thought of as a pyramid (Fig. 1). A broad bottom represents the many basic sciences that enable biomedical science, and a narrow top represents a specific clinical or biomedical problem. If the width of the base is too narrow, then the height of the pyramid will be limited. Solving advanced biomedical problems requires a sufficiently broad base in the fundamental basic sciences. A large edifice cannot be built on a small foundation.

In the United States today, the base of the pyramid is too small. The budget of the National Science Foundation (NSF) — the primary funding agency for all the basic sciences — is only one-quarter of the size of the budget of the NIH — the primary funding agency for biomedical science — and is growing at only half the rate. The number of research papers published in the biomedical sciences in 1991 was about three times the number published in chemistry, physics and mathematics combined³. Biology has rapidly become the dominant undergraduate scientific specialization. In 1988 at the University of California, San Diego, a typical large state institution, about 15.6% of the undergraduates were enrolled in biological sciences and about 8.5% were in chemistry, physics and mathematics combined. The imbalance doubled in the following ten years: in 1998, the numbers were 23.9% and 6.6%, respectively. Bruce Alberts, president of the National Academy of Sciences, has noted that, unless this trend is reversed soon, the United States will lose its seed crop of students as well as of faculty members who can teach the underpinning sciences to biologists⁴.

The importance of nurturing the underpinning sciences is widely recognized. As one

Table 1 Tools that have helped to advance biomedical science

Physics:	Chemistry:	Mathematics and computer science:
X-ray crystallography, synchrotron radiation sources, neutron sources, nuclear magnetic resonance and other spectroscopies for molecular structure determination, MRI, positron emission imaging, microscopies, lasers, molecular tweezers and single-molecule spectroscopies.	Combinatorial and solid-phase syntheses, DNA sequencing, biocompatible materials, drug-delivery devices, polymerase chain reaction, separation and purification methods, and gene chips.	Fast Fourier transforms used in spectroscopy and in CAT scans, fast sequence alignment and database methods used in genomics, conformational search and optimization methods used in protein folding, and ecological and population models of disease.

of the four cornerstones for the future of the NIH, its director, Harold Varmus, has proposed that the agency should “harness the allied disciplines”⁵, and a bioengineering consortium (BECON) has already been initiated. Similar calls have come from Rita Colwell, director of the NSF; from Neal Lane, President Bill Clinton’s science adviser; and from the Federation of American Societies for Experimental Biology in a document called *Molecular Medicine 2020*. Several academic institutions have recognized the need by building large interdisciplinary institutes^{6,7}.

Respecting scientific cultures

But if we recruit allied disciplines, we must accept their standards, methods and timeframes. For example, physicists, engineers, computer scientists and mathematicians working on the development of computational models have certain standards of rigour, with attention paid to the numbers of parameters per prediction, measures of uncertainty, and systematic approximations. Such standards are essential if we are to progress towards a base of knowledge that rests on a well-understood hierarchy of concepts, models and assumptions. Bolder models that pay less attention to standards may appear attractive, but can turn out to be expensive dead ends. Research money will be better spent in the long run if the foundations are strong enough to support the tops of our pyramids.

Harnessing the allied disciplines will not be easy. The pyramid metaphor implies a base that is broad. The underpinning research can be distant from the biomedical payoffs at the top of the pyramid. And, unlike biomedicine, some of it cannot easily capture the public’s imagination and compassion.

Developing tools, foundations, principles and models can lead to big payoffs, but progress is often slow. NMR was originated by I. I. Rabi in 1938, combined with Fourier transform methods by R. R. Ernst and W. A. Anderson in 1966, and first used to obtain a protein structure in 1985. Now, 60 years after its discovery, it is a central tool of structural biology⁸. Combinatorial chemistry, developed by H. M. Geysen in 1984 and now central to industrial pharmaceutical chemistry, has its roots in the solid-phase peptide synthesis developed by R. B. Merrifield in 1963. The all-atom and lattice models that are popular in computational biology have their roots in D. H. Andrews’ 1930 vibrational spectroscopy models, and in polymer theories of the 1940s. Medical radioisotopes emerged from the nuclear physics of the 1940s. From 1992 to 1996, the fraction of protein structures determined using X-rays from synchrotrons rose from 15% to 40%, and there are now year-long waiting lists for synchrotron beam lines. Yet the principle

that X-rays could be obtained from synchrotrons dates back to 1947.

The classic case for long-term funding is X-ray crystallography, arguably the single tool most responsible for the revolution in structural biology and drug design. X-ray diffraction, which was first developed by Max von Laue in 1912, and first applied to protein crystals in 1934, did not lead to a protein structure until 1959. In reviewing the history of the Laboratory of Molecular Biology in Cambridge, England, Max Perutz has written that biomolecular crystallography, which has led to at least a dozen Nobel prizes, could not have arisen under short-term funding mechanisms.

Many other key advances in structural biology, including electron microscopy, atomic force microscopy and molecular tweezers, would also have been unlikely candidates — in their early stages — for short-term funding. According to Perutz, the Laboratory of Molecular Biology “continues to solve problems that would have been considered beyond the reach of science only a short time ago. Some of these, such as Nigel Unwin’s structure of the acetylcholine receptor, or Richard Henderson’s atomic model of bacteriorhodopsin, or John Walker’s mitochondrial ATPase structure, have required sustained efforts lasting many years and could never have been solved if we had had to depend on short-term grants. Best of all, some of our work is now finding applications in practical medicine.”⁹

The Four Horsemen of the Apocalypse that kill US biomedical research grant proposals are: ‘few preliminary results’, ‘slow progress’, ‘few publications’ and ‘too little immediate biomedical relevance’. But these four criteria are precisely the obstacles that often limit our ability to solve the hard problems in the enabling sciences. Sometimes there are no short-cuts. John Sulston, head of the Sanger DNA Sequencing Centre in Cambridge, England, was told in the early 1990s that he was “nuts” to spend millions of dollars “on something [sequencing *Caenorhabditis elegans*] which didn’t solve biological problems right off”¹⁰. Now, *C. elegans* is revealing the molecular secrets of ageing — an ample return on the long-term investment. Similarly, developers of all-atom computational models in biology have sometimes said that US federal funding is readily available for applying models to biological problems, but not for improving the models themselves. Laying the groundwork can be expensive and slow, particularly when compared to the white-hot pace of modern biomedical research.

There are many sound reasons to support the basic sciences. One reason is to help cure disease. But, from biomedicine’s perspective, how do we know when the base of the pyramid is too broad, or when research is so speculative that it will never pay off for bio-

medicine? These judgements are no different from those needed for any other scientific review. Such decisions can be made by enlisting the best people in the field, and charging them with maintaining the appropriate level of vision. In the same way that biologists know when to trust yeast as a model for human genetics or dogs as a model for human cardiology, quantum chemists know what basis sets to trust for chemical geometries, and physicists know what levels of atomic detail are needed in computer models of drug-receptor interactions. In harnessing the allied disciplines, such experts should constitute new study sections or review panels.

In short, we need substantial new funding for the underpinning sciences. Support should go to agencies such as the NIH, NSF and US Department of Energy. Because of the size of this undertaking, the NIH must be centrally involved, possibly with new initial review groups and study sections devoted to the underpinning sciences.

Support long-term research

We also need new review philosophies, mechanisms and guidelines that support long-term scientific studies. I am not suggesting losing focus on the biomedical relevance of research. I am suggesting that we identify the hard problems of basic science that underpin biomedicine, and establish ways to solve those problems. Current grants focus on single investigators solving single problems having payoffs to biomedical research within 3–5 years. Perhaps some research should be supported in a more ‘mission-oriented’ way, in which an investigator solves a small part of a bigger puzzle that is identified as the ‘mission’, with payoffs to biomedicine over longer periods.

New opportunities have sprung up all around us for accelerating the assault on disease. The underpinning sciences have helped create these opportunities, but they are also the source of some of our greatest challenges. Now is the time to strengthen the foundations to seize these opportunities. □

Ken A. Dill is in the Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, USA.

1. Pennisi, E. *Science* **279**, 978–979 (1998).
2. Studt, T. *Drug Discovery and Development* 62–63 (January 1999).
3. Stephan, P. E. *J. Econ. Lit.* **34**, 1199–1235 (1996).
4. Alberts, B. *Cell* **92**, 291–294 (1998).
5. Varmus, H. AAAS Plenary Lecture (13 February 1998).
6. Garwin, L. *Nature* **397**, 3 (1999).
7. Service, R. F. *Science* **284**, 80–83 (1999).
8. Borman, S., Dagani, R., Rawls, R. L. & Zurer, P. S. *Chem. Eng. News* **76**, 39–75 (1998).
9. Perutz, M. <http://www.nobel.se/essays/perutz/index.html>
10. Pennisi, E. *Science* **282**, 1972–1974 (1998).

Acknowledgements This text is modified from a document written on behalf of the US Biophysical Society when the author was president, 1998–99. Thanks to Sarina Bromberg, Jim Cassatt, Ro Kampman, Peter Kollman, I. D. Kuntz, Alex McPherson, Tom Poulos, Jolanda Schreurs, Madeline Shea, Pete von Hippel, Keith Yamamoto and members of the society’s council for comments.

Speciation without isolation

Tom Tregenza and Roger K. Butlin

There is abundant evidence that new species can arise when a population of organisms is split into isolated elements. The occurrence of sympatric speciation — speciation without isolation — is much more contentious but is now set on firmer theoretical ground.

The rise of biomathematics, which led John Maynard Smith to say, “if you can’t stand algebra keep out of evolutionary biology”¹, has been a runaway success. In many fields, empiricists continually struggle to keep up with and verify the assumptions and predictions of modellers. An exception is the famously contentious topic of sympatric speciation — the process by which new species arise from coexisting as opposed to geographically isolated populations. There is growing evidence, particular-

ly from lakes full of closely related fish species², that sympatric speciation does occur in nature. But models of the phenomenon have stubbornly concluded that evolution of sexual isolation without spatial isolation seems very unlikely.

At first glance, sympatric speciation looks straightforward. If a lake contains two potential resources — say, large or small prey — then large or small predatory fish will do well while medium-sized fish will be at a disadvantage. This disadvantage to intermedi-

ates, termed ‘disruptive selection’, creates pressure for divergence into two populations of distinct ecological types.

In sexual populations, the stumbling block preventing sympatric speciation is that mating between divergent ecotypes constantly scrambles gene combinations, creating organisms with intermediate phenotypes (physical characteristics). This mixing can be prevented only if there is assortative mating, in which pairings between similar individuals are more common. With disruptive selection, this pairing pattern is selectively favoured, because it reduces the production of offspring that are less well adapted to their environment. But there is a barrier to the evolution of assortative mating — recombination, the shuffling of genes during gamete formation, which means that genes for mating preference and ecotype (size for instance) may get mixed up whenever an occasional mating between different types occurs. This creates individuals with a preference for the opposite ecotype, increasing gene flow between types and opposing speciation.

Modellers have sought to duck this problem by assuming one of two things — either that the genes responsible for ecological traits and for mating preferences are so close together on a single chromosome that they are only rarely mixed up by recombination³, or that a single gene could both code for the trait and create a preference for partners with that trait. These are plausible assumptions for some situations, such as an insect shifting to a new host plant where it also mates⁴. But in general they are not.

Two new theoretical treatments by Kondrashov and Kondrashov⁵ (KK, page 351 of this issue) and Dieckmann and Doebeli⁶ (DD, page 354) address these difficulties. Both present models in which there are several separate genetic loci for ecological traits and mate preferences (see Box 1 overleaf for details). With these more realistic assumptions, both predict that sympatric speciation can occur without very strong selection against intermediate forms.

There are two variants of the KK model. In the first, mating probability depends on how similar two individuals are for a single marker trait (such as colour); in the second, it depends on a match between male trait and female preference. In the DD approach, mating probability is determined by either the ecological trait or a marker trait, with these loci exerting their influence through a separate set of mating loci. In all of the models, selection increases associations between ecological and marker traits, leading to sexual isolation between ecologically distinct populations (Fig. 1).

In the KK model, disruptive selection is assumed to favour the most extreme phenotypes, regardless of their absolute values. However, this does not fit with the example they give of selection due to two distinct

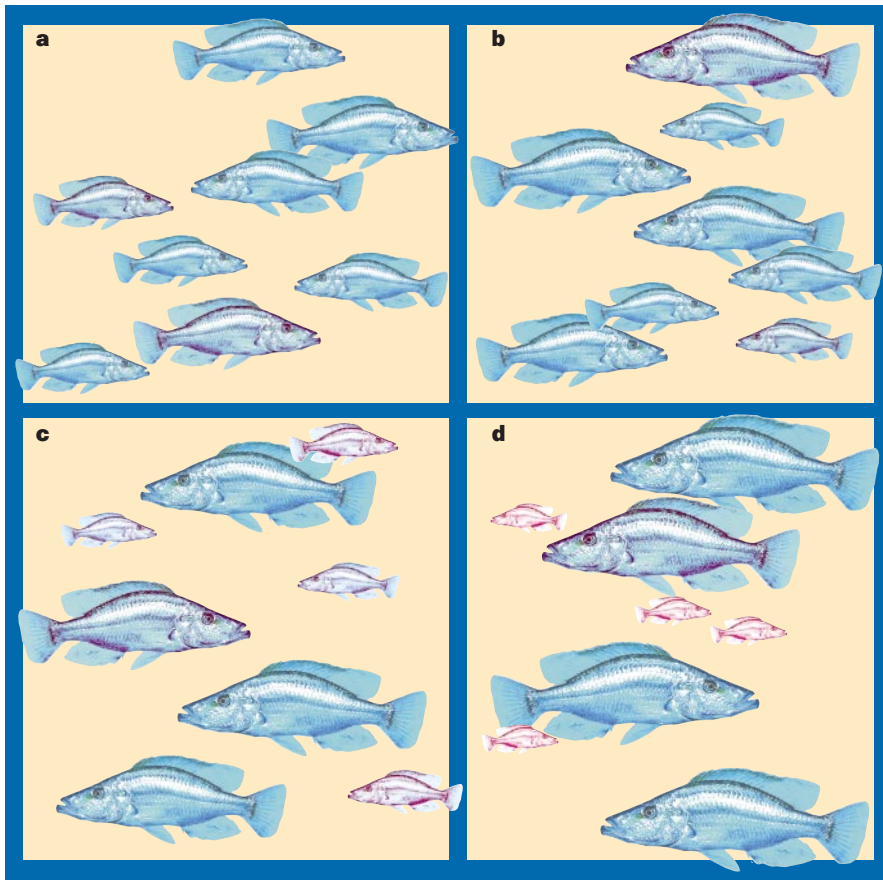


Figure 1 Sympatric speciation according to the KK model⁵. The population is initially uniform for both ecological traits (size) and mating traits (colour), but contains genetic variation for both traits. **a**, The population rapidly evolves to intermediate size. **b**, Disruptive selection creates divergence in ecological traits. **c**, **d**, This in turn creates selection for assortative mating, increasing variation in colour, and increasing association between colour and size. The process is similar in the DD model⁶, except there is greater initial variance in colour, and colour and mating traits diverge at the same time. Unlike previous models, positive covariance between preference and traits develops rapidly despite maximum recombination between genetic loci. The more realistic assumptions^{5,6} about genetic architecture increase the theoretical likelihood that sympatric speciation can occur.

Box 1: Genetic architecture and modelling approach

Although the models discussed here^{5,6} agree on the headline prediction that disruptive selection can drive sexual isolation, and result in sympatric speciation, they differ fundamentally in how they arrive at this conclusion. The DD model allows stochastic variation in genetic composition to create associations between preferences and traits. By contrast, the KK model is deterministic, and simply assumes an initial low level of linkage disequilibrium – a greater-than-chance tendency for preferences and traits to be inherited together.

This difference changes the models' predictions about the effect of the number of genetic loci underlying each trait.

In the DD model, the effect of number of loci is considered in terms of its effect on genetic drift. If more loci are involved, speciation tends to take longer because drift is weakened, giving fewer opportunities for chance increases in association between ecological and mating traits to trigger selection for assortative mating. In the KK model, the parameter considered is the strength of disruptive selection required to drive divergence. In this

case, larger numbers of ecological loci increase the chances of speciation because they increase the production of disfavoured intermediate phenotypes. The numbers of marker loci have the opposite effect, as fewer marker loci mean more extreme (and hence isolated) marker phenotypes. These differences make a direct comparison between the models' predictions impossible, which is a pity: it is important to know how robust the predictions of the models are because they point to ways of identifying cases of sympatric speciation. **T.T. & R.K.B.**

resources. For instance, if a lake contains large and small food items, and fish are initially all small (as in the KK model), then slightly larger individuals will be less efficient at eating small prey, but still hardly any better at eating large prey. Therefore, with two resources, selection does not simply favour the smallest and the largest individuals. It will favour both extremes only if the population begins at an intermediate size, or if there is competition between individuals.

In the DD model, disruptive selection explicitly arises from competition for a single resource — a potentially more common ecological situation. The resource is assumed to have a unimodal distribution such that its carrying capacity is highest when the entire population has a particular phenotype. Selection leads to the phenotype of all individuals initially converging on this point. This provides an explanation for the nagging problem in other models of how the initial population comes to be in a state in which all phenotypes are intermediate and adaptation to the environment is suboptimal. As the population converges on the resource maximum, competition between similar individuals creates selection in favour of those with a divergent genetic make-up that use slightly less abundant resources but experience a more than compensatory decrease in competition. With assortative mating, this process eventually leads to 'evolutionary branching'⁷ — two distinct and reproductively isolated phenotypes are selected despite the resource being unimodal.

An assumption of disruptive selection generated by competition is that there are no

other species that might tend to counter divergence within the focal species through increased interspecific competition. If an absence of other competitors is indeed important, sympatric speciation resulting from competition may be most likely in new, empty habitats. This fits with examples from crater lakes⁸; these lakes were initially empty, but appear to have been colonized by a single

species of fish which has subsequently given rise to a number of species.

As well as being likely to convince many sceptics of the theoretical tractability of sympatric speciation, these new models apply equally to the other contentious example of selection leading to speciation. 'Reinforcement' is an increase in pre-mating isolation between two divergent groups, resulting from selection against hybrid offspring because they are less viable or fertile than their parents⁹. As with sympatric speciation, the main objection to reinforcement has been that recombination will break down the link between mating preferences and genes responsible for hybrid dysfunction. The fact that sympatric speciation can occur despite such recombination indicates that reinforcement may also have escaped its theoretical straitjacket¹⁰. □

Tom Tregenza and Roger K. Butlin are in the Ecology and Evolution Programme, School of Biology, University of Leeds, Leeds LS2 9JT, UK.
e-mail: gentbr@gps.leeds.ac.uk

1. Maynard Smith, J. *Evolutionary Genetics* (Oxford Univ. Press, Oxford, 1989).
2. McCune, A. R. & Lovejoy, N. R. in *Endless Forms: Species and Speciation* (eds Howard, D. J. & Berlocher, S. H.) 291–308 (Oxford Univ. Press, New York, 1998).
3. Felsenstein, J. *Evolution* **35**, 124–138 (1981).
4. Feder, J. L. in *Endless Forms: Species and Speciation* (eds Howard, D. J. & Berlocher, S. H.) 291–308 (Oxford Univ. Press, New York, 1998).
5. Kondrashov, A. S. & Kondrashov, F. A. *Nature* **400**, 351–354 (1999).
6. Dieckmann, U. & Doebeli, M. *Nature* **400**, 354–357 (1999).
7. Geritz, S. A. H., Kisdi, E., Meszena, G. & Metz, J. A. J. *Evol. Ecol.* **12**, 35–57 (1998).
8. Schliewen, U. K., Tautz, D. & Pääbo, S. *Nature* **368**, 629–632 (1994).
9. Butlin, R. K. & Tregenza, T. *Nature* **387**, 551–552 (1997).
10. Kirkpatrick, M. & Servedio, M. R. *Genetics* **151**, 865–884 (1999).

Materials science

Superhard ceramics

R. J. Brook

In contrast to the polymer scientist, who has the privilege of calling upon an ever-expanding array of compositions from which to choose, the ceramic scientist has had to devote years of intensive research to the refinement of a relatively limited number of systems. So it is exciting news to hear from Zerr *et al.* on page 340 of this issue¹ that in one of these systems — silicon nitride — a new form of compound has been found. Moreover, the new cubic compound is potentially much harder than existing phases, offering considerable industrial scope for the material.

Silicon nitride² (Si₃N₄) has two long-established crystal forms, α and β . In both, the central silicon is linked to four surrounding nitrogens in a tetrahedral array (Fig. 1a). The different crystal structures are then distinguished by the ways in which the set of tetrahedra are linked to one another. In Si₃N₄, each tetrahedron is linked at each corner to

two others, conferring a greater degree of structural rigidity than for silicate systems where the tetrahedra (formed from silicon and four oxygens) are linked one-to-one at the corners.

In the ceramics community, Si₃N₄ has enjoyed an exceptional degree of attention largely because of its potential as a material for high-temperature heat engines. The excitement arises in part from its mechanical properties, such as greater strength and ability to resist mechanical failure when subjected to sudden temperature changes. It is also less brittle than many ceramics, owing to the way in which the crystalline grains in the material become acicular (that is, length ten times greater than width); the resulting interlinking improves the resistance to mechanical failure. Ceramics are traditionally prepared from powdered materials, and Si₃N₄ is no exception: the precursor material is shaped into a powder mass, which is then

consolidated by heat treatment. A special aspect of Si_3N_4 processing is the use of both phase types: when α powder is used, it converts during heat treatment to the β form to yield the acicular crystal morphology.

The application in which Si_3N_4 has been most successfully exploited is for cutting tools which make use of its hardness and resilience. In this area, attention has turned to the so-called sialons³, which are based on silicon nitride (both α and β types) but in which substitutions of aluminium and oxygen are made respectively on the silicon and nitrogen sublattices. As well as providing a basis for wide-ranging and subtle use of additive elements, the sialons have opened doors to a greater range of properties, including improved hardness and resistance to corrosion.

The original objective of using Si_3N_4 to produce engine components for operation at high temperatures and pressures has not yet been reached. Despite the 'ceramic fever', during which intensive international effort was expended for more than a decade, the current high-profile use of Si_3N_4 in engines is limited to valves. Moreover, the use of Si_3N_4 in failure-critical components in aircraft continues to be ruled out by their susceptibility to brittle fracture, namely the risk of sudden and total failure under impact as a result of some small — fractions of a mil-

limetre — processing flaw within a component.

So what of the new form that has been identified by Zerr *et al.*¹? The 'spinel' form of Si_3N_4 that results from high-pressure fabrication (15 gigapascals at temperatures over 2,000 K) offers several avenues of interest. The first is the structure (Fig. 1b). In the spinel system, the nitrogen anions adopt a close-packed, cubic crystal structure, and the silicon cations occupy spaces in the lattice between the anions. The silicon cations are arranged in an unusual two-sublattice system, with two-thirds of the cations surrounded by six anions (an octahedral array) and one-third surrounded by four anions (a tetrahedral array).

In contrast to the other forms of Si_3N_4 , the structure of cubic Si_3N_4 is more dense, stiffer and with every expectation of being considerably harder, perhaps even as hard as stishovite (the third hardest material after diamond and cubic boron nitride). Most surprising is the existence of silicon in an octahedral sublattice. The higher-density spinel structure results from using high pressures during fabrication, as predicted by Le Chatelier's principle, which says that if a system in chemical equilibrium is disturbed, it will react in a way that tries to neutralize the disturbance and restore the equilibrium. But a key benefit is that, once the pressure is removed, the material survives in a metastable state and is, indeed, stable to at least 700 K.

In areas where hardness and wear resistance are required (such as cutting tools), the new spinel structure offers advantages over current Si_3N_4 phases. The work by Zerr *et al.*¹ also demonstrates particularly well the trend in materials science, where experimental studies and theoretical studies march closely in step and with mutual benefit. Their theoretical calculations throw light both on why the high-density form of Si_3N_4 should be spinel and on the stiffness and hardness of the new material.

The quest for ever more reliable processing methods has had several consequences. In seeking to avoid the presence of any flaws, greater attention has been given to precursor materials that are themselves free of flaws. This demands new techniques for producing advanced ceramics. Sol-gel processing for oxides⁴, in which small particles in a liquid suspension are converted into an amorphous solid, and polymer precursors for non-oxides⁵ are just two examples. It can be expected that following these avenues will lead to a remarkable array of metastable intermediate structures. Heat treatment of polymer precursors, for example, provides an accessible route to compounds containing silicon, nitrogen, carbon or boron, where amorphous structures of remarkable resilience have already been identified. In this sense, the finding of the new cubic sili-

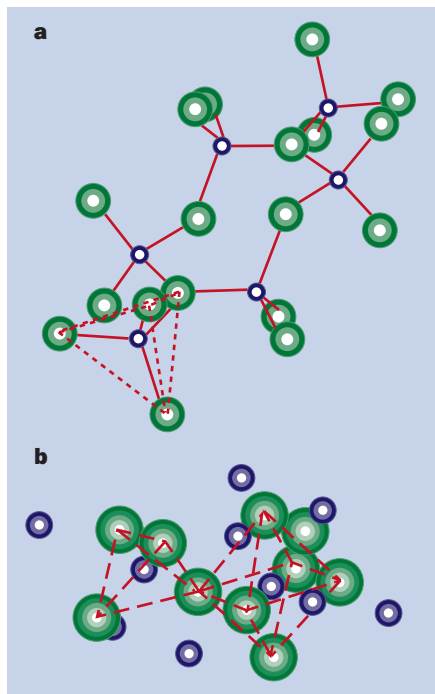


Figure 1 Phases of silicon nitride, showing the silicon atoms in purple and the nitrogen atoms in green. a, In the β form of Si_3N_4 , the relatively open packing of the nitrogen results from the corner-sharing tetrahedra, one of which is outlined. b, In the new cubic form of Si_3N_4 created by Zerr *et al.*¹, the nitrogen is close packed and silicon occurs in both tetrahedral (shown on the left) and octahedral (shown on the right) surroundings.



100 YEARS AGO

So far as the details of practical cultivation are concerned, we are not so much in advance of our forefathers. We have infinitely greater advantages, and we have made use of them, but if they had had them they would have done the same. We are able to bring to bear on our art not only the "resources of civilization" to a degree impossible to our predecessors, but we can avail ourselves also of the teachings of science and endeavour to apply them for the benefit of practical gardening. We are mere infants in this matter at present, and we can only dimly perceive the enormous strides that gardening will make when more fully guided and directed by scientific investigations. One object of this conference is to show that cultural excellence by itself will not secure progress, and to forward this progress by discussing the subject of cross-breeding and hybridisation in all their degrees, alike in their practical and in their scientific aspects. To appreciate the importance of cross-breeding and hybridisation we have only to look round our gardens and our exhibition-tents, or to scan the catalogues of our nurserymen.

From *Nature* 20 July 1899.

50 YEARS AGO

To find a principle common to a science of physics which, according to some thinkers, is nearing completion, and a science of psychology which, according to all, has scarcely begun, is inevitably to put oneself in the company of the dialectical materialists, the kinematical relativists and others, whose universal, fundamental principles of Nature are alike seductive and baseless and are further alike in being all incompatible with one another. If one likes that kind of thing, one can take a choice between the principle that all things generate their contraries with which they then identify themselves, the principle that all experience must be expressible in terms of intuitive apprehension of the passage of time ... and any other principle that one can think of. I know no way of choosing between them; it seems to depend on which siren's voice sounds the sweetest. I can only recommend, in the strongest possible terms, a liberal use of wax and straps.

From *Nature* 23 July 1949.

con nitride is part of a wider search in which the array of classical ceramic compounds is being expanded. □

R. J. Brook is at the Engineering and Physical Sciences Research Council, Swindon SN2 1ET, UK.

1. Zerr, A. *et al. Nature* **400**, 340–342 (1999).

2. Hampshire, S. *Mater. Sci. Technol.* **11**, 119–171 (VCH, Weinheim, 1994).
 3. Jack, K. H. in *New Oxide Technical and Engineering Ceramics* (ed. Hampshire, S.) 1–30 (Elsevier, London, 1986).
 4. Segal, D. *Mater. Sci. Technol.* **17A**, 70–98 (VCH, Weinheim, 1996).
 5. Riedel, R. *Mater. Sci. Technol.* **17B**, 1–50 (VCH, Weinheim, 1996).

Neurobiology

Discriminating migrations

Pasko Rakic

Many brain structures, like some frontier countries, are composed exclusively of immigrants. In the mammalian brain, for example, virtually all neurons migrate from their origin, near the cerebral ventricles, to the distant territories where they establish permanent residence. Although the mechanisms that attract neurons to their final destinations are not well understood, we know that certain brain structures discriminate between cell populations by discouraging specific classes of neuron from invading their territories. On page 331 of this issue, Wu *et al.*¹ reveal a molecular cue that guides neurons as they migrate from the postnatal rodent forebrain into the olfactory bulb. They show that this guidance mechanism depends on the brain structures that surround the migratory pathway — structures that repel migrating neurons from their borders, while leaving invasion of the olfactory bulb as an option.

In the developing nervous system, most neurons are generated at different sites from those in which they permanently reside². The intervening process, termed neuronal migration, denotes the displacement of a neuronal cell body from its origin in the proliferative zones to its final destination in the mature brain. After they have finished dividing, cells dislodge from adjacent progenitors, extend a 'leading process', then move along specific pathways³. Interactions between the migrating neurons and the surfaces of neighbouring cells are crucial for the selection of migratory pathways, as well as for orientating, conducting and stopping neuronal movement at the appropriate site³.

The main difference between neuronal migration and axonal navigation is movement of the cell body. Whereas it undergoes a unidirectional translocation in neuronal migration, the cell body remains stationary in axonal movement. Translocation is regulated by specific ion channels/receptors, which cooperatively regulate the calcium influx that is essential for the cytoskeletal rearrangements affecting the rate at which the neuron moves⁴. With regard to pathway selection, migrating neurons fall into two main categories — gliophilic cells, which show an affinity for the elongated radial glial fibres that guide their movement, and neurophilic cells, which preferentially stick to the

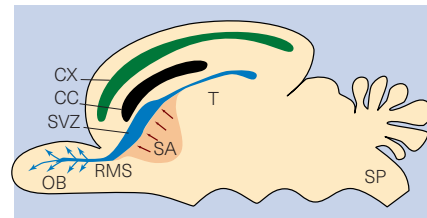


Figure 1 Possible function of Slit in directing neuronal migration in the postnatal mouse brain. The blue colour indicates migration of neurons from their origin in the subventricular zone (SVZ), situated at the surface of anterior cerebral ventricles, to the olfactory bulb (OB) via the rostral migratory stream (RMS). Red arrows indicate the presence and direction of repulsive influences — one of which has been identified by Wu *et al.*¹ as Slit — that emanate from the surrounding structures of the forebrain such as the septal area (SA), CX, cortex; CC, cerebellum; CC, corpus callosum; T, thalamus; SP, spinal cord.

surface of other neurons. The neurophilic cells usually move tangentially, bypassing radial glial shafts², and a good example of this is translocation of cells from the anterior portion of the telencephalic subventricular zone to the olfactory bulb⁵. The olfactory bulb is unique in that it continues to import these neurons even in the mature brain⁶, and the newly generated neurons move tangentially as a cohort of closely apposed cells called the rostral migratory stream.

Why do neurons generated in the telencephalic subventricular zone migrate to the relatively distant olfactory bulb, rather than to nearby structures such as the septum or anterior forebrain? The reason was suspected to be selective recognition of a particular substrate by the migratory neurons. Hu and Rutishauser⁷ initially suggested that a repulsive molecule in the septum could prevent neurons from the subventricular zone from entering septal territory. Wu *et al.*¹ now propose a specific protein that does this job.

This protein turns out to be a member of the Slit family of proteins, which were initially identified in the fruitfly *Drosophila melanogaster*⁸ and have since been found in the brains of amphibians, birds, rodents and primates — including humans. These proteins rose to prominence when they were found to be important in controlling axon

guidance. Then, earlier this year, four papers (reviewed in ref. 9) showed that an interaction between the Slit proteins and a cell-surface receptor called Roundabout (Robo) is involved in preventing axons from crossing the embryonic midline.

Wu *et al.* now suggest a previously unsuspected function for Slit in directing neuronal migration, well before the initiation of axonal connections (Fig. 1). By co-culturing tissue explants from various brain structures, the authors confirmed that a repulsive activity acts on migrating neurons in the brain. They also showed that Slit is expressed in the postnatal septum, and that it is chemorepulsive for migration of subventricular-zone neurons from explants, as well as for migration of these neurons within the rostral migratory stream. Although this effect seems to be repulsive, it does not inhibit the migration, and it depends on a concentration gradient of the Slit protein. By blocking endogenous Slit using the extracellular domain of Robo, which is a receptor for Slit, Wu *et al.* showed that Slit contributes to the repulsive activity of the septum for neurons of the subventricular zone.

The novelty of Wu and colleagues' findings lies in the fact that the Slit molecules are diffusible, distributed as a concentration gradient that can be read by migrating cells with the appropriate receptor. In a parallel study in *Neuron*¹⁰, the same group shows that Slit also acts as a guidance cue for neurons migrating from the ganglionic eminence to the neocortex in the embryonic brain. Anderson *et al.*¹¹ previously found that a selective class of neurons from the ganglionic eminence does not migrate radially, but, rather, that these neurons invade the subventricular zone tangentially on their way to different cortical areas. Now it seems that Slit, which is present in the ventricular zone of the ganglionic eminence, directs the migration of cells from this region to the cortex¹⁰.

So, although repulsive molecules were thought to guide the formation of axonal connections, Wu *et al.* provide the first experimental evidence that a similar mechanism directs neuronal cell migration. Their study is also an example of how new approaches in experimental biology allow us to examine the roles of individual genes and their products in complex, multicellular developmental events, and to look at their involvement in congenital malformations. The findings also provoke new ideas for therapeutic strategies. For instance, repulsive molecules might be used to prevent metastasis in the brain, or to streamline transplanted cells into parts of the brain affected by neurodegenerative disorders.

Motility is one of the primordial features of prokaryotic and eukaryotic cells. From the beginning of animal evolution, mobile cells were adapted to move within the environment — either towards positive chemical stimuli or away from negative ones. This

basic strategy seems to be preserved even in nerve cells. It is not, of course, enough to place neurons in their final position in the brain, because this process also requires surface-mediated interactions between migrating neurons and adjacent cells^{3,12}. But amplification of the ancient strategies that operate in the most primitive single-celled organisms may be combined with the newer developmental mechanisms in a cooperative attempt to build a complex nervous system.

Pasko Rakic is in the Section of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06510, USA.
e-mail: pasko.rakic@yale.edu

1. Wu, W. *et al.* *Nature* **400**, 331–336 (1999).
2. Rakic, P. *Experientia* **46**, 882–891 (1990).
3. Pearlman, A. L. *et al.* *Curr. Opin. Neurobiol.* **8**, 44–54 (1998).
4. Komuro, H. & Rakic, P. *J. Neurobiol.* **37**, 110–130 (1998).
5. Lois, C., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. *Science* **271**, 978–981 (1996).
6. Luskin, M. B. *J. Neurobiol.* **36**, 221–233 (1998).
7. Hu, H. & Rutishauser, U. *Neuron* **16**, 933–940 (1996).
8. Nüsslein-Vollhard, C., Wieschaus, E. & Kluding, H. *Wilhelm Roux Arch. Dev. Biol.* **193**, 267–283 (1988).
9. Harris, W. A. & Holt, C. E. *Nature* **398**, 462–463 (1999).
10. Zhu, Y., Li, H., Zhou, L., Wu, J. Y. & Rao, Y. *Neuron* **23**, 473–485 (1999).
11. Anderson, S. A., Eisenstat, D. D., Shi, L. & Rubenstein, J. L. R. *Science* **278**, 474–476 (1997).
12. Rakic, P., Cameron, R. S. & Komuro, H. *Curr. Opin. Neurobiol.* **4**, 63–69 (1994).

Astronomy

Big planets and little stars

James Liebert and William B. Hubbard

The unambiguous discoveries of giant, extrasolar planets and very cool 'brown dwarf' stars have only been made in the past few years. Perhaps it was the dramatic detection of the companion brown dwarf star, Gliese 229B — which had methane in its spectrum and evidence for dust and cloud formation — that has most linked the communities of astronomers interested in giant planets and cool stars, who met together for the first time last month*. If Gliese 229B were a true star, then its atmosphere would be so much hotter that neither methane nor most types of clouds could form. Astronomers have always categorized stars according to their temperature: the brightest and hottest are O stars, followed by B, A, F, G stars (which include our Sun), K and, at the bottom of the scale, M stars, which rarely reach a cool 2,500 K.

These are exciting times, with fresh discoveries of new classes of cool astrophysical objects, many of which have masses too low to fuse hydrogen into helium — the definition of a star. The detection of stand-alone brown dwarf stars with temperatures similar to Gliese 229B near 1,000 K led to a new category of 'methane' or 'T dwarf' stars. These have infrared spectra dominated by methane and gaseous water, more like Jupiter than a small M dwarf star. The Sloan Digital Sky Survey (SDSS) reported the discovery of the first such object without a companion star the week before this meeting¹. Stunning infrared and optical spectra of this and a second SDSS discovery were shown at the meeting (J. Pier, US Naval Observatory, Arizona). Hot on the heels of Sloan, however, is the brown dwarf group working on infrared sources found in the Two Micron All Sky Survey (2MASS), which have spectra of four T dwarfs (A. Burgasser, Caltech).

The 2MASS group has discovered nearly one hundred of the warmer 'L dwarf' stars, another category of substellar objects that are still below M dwarf temperatures. They have modelled the observed L dwarf properties using a theory of stellar evolution calculated by A. Burrows and colleagues (Univ. Arizona). Using this model of the L dwarf data it is possible to estimate the likely spatial density and luminosities of substellar objects (N. Reid, Caltech). The 2MASS group predicts that most L dwarf stars are cooler and less luminous than Gliese 229B, with temperatures down to several hundred degrees kelvin (or less). It is possible that these cool dwarfs are as numerous as all the stars in our local neighbourhood. In contrast, a low-resolution infrared spectrum of a possible planetary-mass object, previously reported elsewhere², now appears to show evidence for carbon monoxide, which indicates a temperature closer to 3,000 K (in the range of cool stars) rather than 1,600 K.

Many L dwarfs appear to be rotating, with velocities up to 80 km s⁻¹ (E. Martin and G. Basri, Univ. California, Berkeley). The rapid rotation of these objects means that they could be very young, which would explain why a substellar object is still a fairly luminous infrared source. Alternatively, they may be old (and therefore stellar) objects, implying that such low-temperature stars are more like Jupiter, which has a short rotation period of about ten hours, than the Sun — a slow rotator that loses its angular momentum because of an internal magnetic dynamo. Substellar objects in the Pleiades cluster and the very young (2–7 Myr) Sigma Orionis cluster have masses lower than 0.075–0.08 of the solar mass, below which theoretical calculations predict hydrogen fusion cannot take place (R. Rebolo, Instituto de Astrofísica de Canarias). The objects in Pleiades extend down to 40 Jupiter masses (0.04 solar mass)

or less; in Orion, substellar objects may have been found near 15 Jupiter masses (0.015 solar). The discovery of such low-mass objects indicates that the formation of a cluster of stars can also produce objects truly intermediate between the lowest-mass stars and the mass of Jupiter, that is not very close in mass to either. This narrows further the distinction between giant planets and cool stars.

The prospects of finding many more extrasolar planets and brown dwarf stars are looking good with a dedicated 1.2-metre telescope at the European Southern Observatory in Chile (D. Queloz, Jet Propulsion Lab., Pasadena). Michel Mayor's group at the Observatoire de Genève is measuring radial velocities of some 1,500 G–K dwarf stars. Although the results to date should be heavily biased in favour of finding massive, close companions, it appears nonetheless that planetary companions less massive than about ten Jupiters are much more common than substellar companions above this mass. That is, brown dwarfs may be more common as stand-alone objects than as companions.

It is not yet clear whether extrasolar planets and brown dwarfs would have a banded appearance similar to Jupiter's (G. Schubert, Univ. California, Los Angeles). According to theory and numerical simulations, these rapidly rotating objects should transport energy to their atmospheres by means of thermal convection from the cooling interior. The style of convection is influenced by the rapid rotation, which leads to cells forming columnar rolls parallel to the rotation axis. The ends of the cells form bands that vary in time, similar to the bands of clouds on Jupiter. The critical parameter is the Rayleigh number, which determines the character of the convection and increases by about seven orders of magnitude from Jupiter to the most massive brown dwarfs. Schubert predicts that brown dwarfs might also have bands, but that the banded structure will be much more chaotic and time-variable than in Jupiter.

Astronomers are beginning to understand the importance of planet-like processes such as meteorology and cloud formation in cool atmospheres. According to thermochemical models, 'clouds' of silicate are buried at high pressures deep below Jupiter's visible atmosphere. These theoretical clouds have now been detected in the atmospheres of L dwarfs at temperatures above 1,000 K, where they absorb more strongly at shorter wavelengths and so redden the emitted near-infrared light. The clouds suddenly disappear in methane dwarfs at effective temperatures cooler than 1,000 K, as predicted by theory. But a sobering account of the complexity of cloud microphysics highlights a basic need

*From *Giant Planets to Cool Stars*, Flagstaff, Arizona, USA, 8–11 June 1999.

for better planetary models in this area (A. Ackerman, Univ. Colorado).

Chemical reactions and phase transitions in the atmospheres and interiors of cool stars and giant planets will complicate the interpretation of what astronomers call 'metallicity' and planetary scientists call departures from solar abundances — that is, spectral evidence for quantities of elements heavier than helium. Results from the Galileo entry probe mass spectrometer³ provide evidence for such departures in Jupiter (D. Hunten, Univ. Arizona), but the additional complication in brown dwarf stars is that some of these objects may be intrinsically non-solar in composition. Broad-brush comparisons between theory and observation are satisfactory in many cases, especially from the perspective of our understanding of lithium depletion by ther-

monuclear reactions in the more massive brown dwarfs, where it acts as a precise gauge of the maximum central temperature reached in these objects. But a full interpretation of the spectra of these cool, tiny, hydrogen-rich masses will eventually require much more elaborate measurements and modelling, because we still don't fully understand interior processes even in Jupiter. □

James Liebert is in the Department of Astronomy, and William B. Hubbard is in the Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721, USA.
e-mails: liebert@stsci.edu; liebert@as.arizona.edu
hubbard@lpl.arizona.edu

1. Strauss, M. *et al. Astrophys. J. Lett.* (in the press); <http://xxx.lanl.gov/abs/astro-ph/9905391>
2. Terebey, S., Van Buren, D., Padgett, D. L., Hancock, T. & Brundage, M. *Astrophys. J. Lett.* **507**, L71 (1998).
3. Niemann, H. B. *et al. J. Geophys. Res.* **103**, 22831–22845 (1998).

Behavioural pharmacology

Breaking the chain of addiction

Gary Aston-Jones and Jonathan Druhan

Two hallmarks of drug addiction are the compulsive nature of substance use and the high propensity for relapse. Relapse is, in fact, so prevalent that attempts to stop using drugs often lead to repeated cycles of detoxification followed by a return to compulsive substance abuse. How can these cycles be broken? A report by Pilla *et al.*¹ (page 371 of this issue) indicates that, for cocaine at least, conditioned cravings and drug-seeking responses may be reduced using a newly developed partial agonist for the dopamine D₃ receptor.

The craving for drugs in an addict is driven not only by the pain and discomfort of withdrawal, but also by the expected rewarding effect of the drug. Craving can even be elicited by stimuli previously associated with taking the drug. Childress, O'Brien and colleagues^{2,3}, for instance, found subjective reports of craving in abstinent opiate or cocaine addicts shown films of the paraphernalia associated with taking drugs. This conditioned craving effect remained strong after years of abstinence, and could be elicited by drug-associated stimuli as apparently innocuous as money. These results illustrate the powerful hold that learned associations have over addicts.

Over the past several years, researchers have begun to work out the neurochemical substrates of such conditioned responses. The brain neurotransmitter dopamine is critical for the 'reinforcing' effects of many drugs of abuse, including cocaine. Cocaine causes large increases in the amount of dopamine available to stimulate dopamine receptors. These increased levels are responsible for the reinforcing effects, meaning that an addict is more likely to seek cocaine in

the future. Consistent with this view, animal studies show that damage to dopamine neurons, or blockade of dopamine receptors, dramatically reduces the animal's motivation to self-administer cocaine⁴. A critical site for this effect is the nucleus accumbens, located in the ventral forebrain⁵. Other studies have shown that, in this region, dopamine is also involved in the conditioned effects of sensory stimuli associated with cocaine⁶.

The dopamine D₃ receptor subtype is thought to be involved in the reinforcing effects of cocaine. This subtype is preferentially located in limbic regions of the brain, which are important in drug reinforcement, and is found at high levels in the nucleus accumbens^{7,8}. So, D₃ receptors could be a target for the dopamine that is critical for cocaine reinforcement. Studies by Caine, Koob and

colleagues⁹ indicated that stimulation of the D₃ receptor reduces self-administration of cocaine by mimicking the effects of the drug. This suggests that an agonist of the receptor could substitute for the abused drug, providing a methadone-like treatment for cocaine abuse. Although not ideal, methadone is certainly useful in treating heroin addiction.

Pilla *et al.*¹ now extend this line of work, presenting new approaches to pharmacotherapies for treating cocaine addiction. They have developed a selective partial agonist for D₃ receptors, called BP 897 (Fig. 1). The authors show that BP 897 acts preferentially at D₃ receptors, and that it has properties of a partial agonist. Unlike a full agonist of D₃ receptors (such as dopamine), a partial agonist has a high affinity but low intrinsic activity at its receptor. This means that BP 897 may act as either a weak agonist or an antagonist, depending on the amount of dopamine at the D₃ receptors. So, at times of low dopamine activity (such as during withdrawal), BP 897 may stimulate D₃ receptors. But the same dose of BP 897 may antagonize D₃ responses stimulated by high dopamine activity (such as in the presence of cocaine).

Partial agonists have considerable potential for treating drug addiction¹⁰. By acting as an agonist in the absence of cocaine, a D₃ partial agonist such as BP 897 may substitute for cocaine and help to offset cravings. In the presence of cocaine (and the associated high concentration of extracellular dopamine), a D₃ partial agonist may prevent the overstimulation of D₃ receptors and blunt the reinforcing effects of cocaine. Also, a partial agonist should have fewer side effects than a conventional dopamine antagonist, owing to its off-setting agonist properties. So, partial agonists may be better tolerated and more likely to be taken than antagonist drugs.

To test BP 897 as a medication for limiting cocaine cravings, Pilla *et al.* used a behavioural model that measures the secondary reinforcing effects of drugs. A drawback of

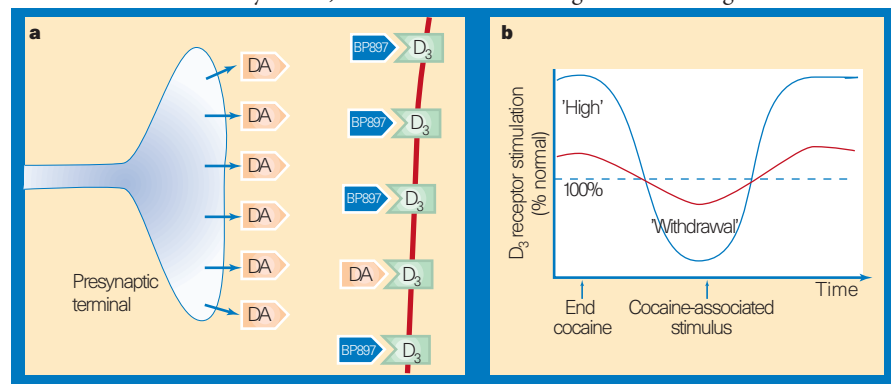


Figure 1 Model of how a partial dopamine agonist (BP 897) may reduce cocaine craving. **a**, Dopamine (DA) is released by a cocaine-associated stimulus. Pilla *et al.*¹ suggest that, by acting as a partial agonist, BP 897 may weakly stimulate dopamine D₃ receptors but block overstimulation of these receptors by high levels of dopamine (such as occurs during administration of cocaine). **b**, BP 897 (red curve) may 'buffer' the changes in D₃-receptor stimulation produced by cocaine or cocaine-associated stimuli, reducing craving during withdrawal. The blue curve represents D₃-receptor stimulation in the absence of BP 897.

conventional self-administration tests for analysing cravings is the fact that, once an animal self-administers its first dose of cocaine, subsequent responses are influenced by response-altering side-effects of cocaine. The authors circumvented this problem by measuring responses to cocaine-associated stimuli during an initial 15-minute period before the first cocaine infusion. In this way, the effects of potential anti-craving medications can be tested without concern for confounding effects of cocaine itself. This behavioural model also directly tests the effects of pharmacotherapies on the conditioned effects of cocaine, known to be important in cocaine abuse.

The authors found that BP 897 selectively reduces the behavioural response for cocaine-related stimuli during the initial 15-minute period, but has no effect on this response after cocaine has been injected. They also showed that BP 897 does not substitute for cocaine in a test of self-administration, nor does it affect the rate at which animals self-administer cocaine. This indicates that BP 897 may not have abuse potential itself — an important attribute. The fact that BP 897 decreases behavioural responses in the initial 15 minutes is particularly exciting. It indicates that BP 897 may reduce the cravings elicited by stimuli previously associated with the drug, without preventing the reinforcement produced by the drug itself or (presumably) by natural reinforcers such as food. The lack of an effect after cocaine administration also indicates that BP 897 may have limited side effects, such as motor or memory disturbances. So, BP 897 may allow the motivated cocaine abuser to abstain, even in the face of stimuli previously associated with cocaine.

Pilla *et al.* provide a potentially powerful new means of treating cocaine addiction, but their findings also raise a number of questions. For example, what actions of the partial D_3 agonist are responsible for the decreased behavioural efficacy of the cocaine-associated cue? Why is its partial-agonist property critical to BP 897's selective effect on conditioned responses? Where in the central nervous system do these actions occur? The authors suggest that the nucleus accumbens may be the critical site, although other regions of the brain also contain D_3 receptors (the amygdala, for instance⁷). What role could BP 897's possible autoreceptor actions — which decrease the release of dopamine — have in these types of behavioural effect? Human dopamine neurons, in contrast to those in the rat, seem to contain few D_3 receptors⁷, so the function of BP 897 as an autoreceptor agonist may differ between the two species. Finally, will BP 897 be effective only as long as an addict stays off cocaine? After all, Pilla *et al.* found that self-administration was not affected by this compound after injection of cocaine.

Studies in humans will be needed to determine the success of partial D_3 agonists such as BP 897 in helping the addict to break the chain of addiction. Whatever the answers, Pilla and colleagues' approach should stimulate the development of a new generation of pharmacotherapies for cocaine addiction. □

Gary Aston-Jones and Jonathan Druhan are in the Department of Psychiatry, University of Pennsylvania School of Medicine, VAMC (151), University and Woodland Avenues, Philadelphia, Pennsylvania 19104, USA.
e-mail: gaj@mail.med.upenn.edu

Mathematics

Randomness everywhere

C. S. Calude and G. J. Chaitin

In a famous lecture in 1900, David Hilbert listed 23 difficult problems he felt deserved the attention of mathematicians in the coming century. His conviction of the solvability of every mathematical problem was a powerful incentive to future generations: "Wir müssen wissen. Wir werden wissen". (We must know. We will know.) Some of these problems were solved quickly, others might never be completed, but all have influenced mathematics. Later, Hilbert highlighted the need to clarify the methods of mathematical reasoning, using a formal system of explicit assumptions, or axioms. Hilbert's vision was the culmination of 2,000 years of mathematics going back to Euclidean geometry. He stipulated that such a formal axiomatic system should be both 'consistent' (free of contradictions) and 'complete' (in that it represents all the truth). Hilbert also argued that any well-posed mathematical problem should be 'decidable', in the sense that there exists a mechanical procedure, a computer program, for deciding whether something is true or not. Of course, the only problem with this inspiring project is that it turned out to be impossible.

In 1931 Kurt Gödel showed that if you assume a formal axiomatic system containing elementary arithmetic is consistent (which is the minimum requirement — if you can prove false results that is embarrassing), then you can prove that it is incomplete. This was a huge surprise; everyone else thought Hilbert was right. The third condition remained open until Alan Turing demolished Hilbert's dream in 1936. Turing showed that no mechanical procedure, and therefore no formal axiomatic theory, can solve Turing's halting problem, the question of whether a given computer program will eventually halt. Turing's argument was based on computable real numbers. A real number, such as π , is a length measured with arbitrary precision, with an infinite number of digits. And a real number is computable if there is a

1. Pilla, M. *et al.* *Nature* **400**, 371–375 (1999).
2. O'Brien, C. P., Childress, A. R., McLellan, A. T. & Ehrman, R. *Ann. NY Acad. Sci.* **654**, 400–415 (1992).
3. Childress, A. R., McLellan, A. T., Ehrman, R. & O'Brien, C. P. *NIDA Res. Monogr. Ser.* **84**, 25–43 (1988).
4. Koob, G. F. & Bloom, F. E. *Science* **242**, 715–723 (1988).
5. Roberts, D. C., Corcoran, M. E. & Fibiger, H. C. *Pharmacol. Biochem. Behav.* **6**, 615–620 (1977).
6. Fontana, D. J., Post, R. M. & Pert, A. *Brain Res.* **629**, 31–39 (1993).
7. Gurevich, E. V. & Joyce, J. N. *Neuropsychopharmacology* **20**, 60–80 (1999).
8. Sokoloff, P., Giros, B., Martres, M.-P., Bouthenet, M.-L. & Schwartz, J.-C. *Nature* **347**, 146–151 (1990).
9. Caine, S. B. *et al.* *Neuroreport* **8**, 2373–2377 (1997).
10. Pulvirenti, L. & Koob, G. F. *Trends Pharmacol. Sci.* **15**, 374–379 (1994).

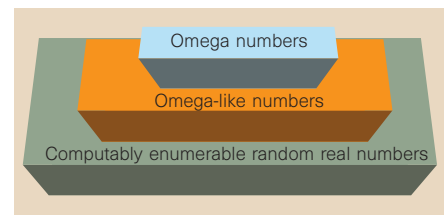


Figure 1 Complexity theorists used to believe that these three classes of real numbers were different. Work by Berkeley mathematician Theodore Slaman² shows that these three classes of real numbers in fact collapse into a single class. This work helps to clarify the nature of randomness in pure mathematics — it is as fundamental as it is in theoretical physics.

computer program or algorithm for calculating its digits one by one. There are programs for calculating π , but it is a surprising fact that nearly all real numbers are not computable. Turing showed that if you could find a mechanical procedure to decide if a computer program will ever halt, then you could compute a real number that is not actually computable, which is impossible.

In 1975 a more modern version of Turing's halting problem emerged. This is Chaitin's Omega number, which is the probability that an arbitrary computer program will eventually halt¹. Omega is computable in a weak sense, but its binary digits or bits are algorithmically random and cannot be distinguished from the result of independent tosses of a fair coin. Now Theodore Slaman² and Robert Solovay³ have greatly increased our understanding of Omega and the depth and pervasiveness of its algorithmic randomness. Their work brings up to date the theoretical context for studying Omega, called algorithmic information theory^{4,5}, and reinforces the limits placed on the power of mathematical reasoning by this theory⁶.

To understand Omega better, let's compare it with another well-known real number, π . The digits of π (3.1415926...) can be

computed one by one; nonetheless, if examined locally, without being aware of their provenance, they appear 'random'. People have calculated π out to one billion or more digits. One of the reasons for doing this, besides breaking the world record, is the question of whether each digit occurs the same number of times. It looks likely, but remains unproven, that the digits 0 through 9 each occur 10% of the time in a decimal expansion of π . If this turns out to be true, then π would be called a simply normal real number. But although π may be random in so far as it's 'normal', it is far from random in the sense of algorithmic information theory, because its infinity of digits can be compressed into a concise computer program for calculating them.

Now let's consider Omega. Although it has a simple mathematical definition, this definition does not enable us to determine more than a finite number of its digits, which can be shown to be 'normal' in base ten and indeed in any base (for example, in binary, 0 and 1 will occur exactly 50% of the time). Moreover, although the infinite amount of information contained in Omega's digits is algorithmically incompressible, it turns out that Omega is 'computably enumerable', which means that it can be calculated by an infinite process during which one can never know how close one is to the final value. In this way, the halting probability that Omega shares two apparently irreconcilable properties: 'algorithmic randomness' and 'computable enumerability'.

An Omega is computably enumerable because a systematic run of all programs will produce better and better approximations (without being able to compute its digits exactly), and random because it is incompressible; there is no better way to find its digits than by tossing a fair coin. No formal mathematical theory can determine more than a finite number of digits of an Omega. In fact, one can explicitly compute a limit on the number of digits of Omega that a specific theory can determine^{4,5}. Berkeley mathematician Robert Solovay³ has now constructed the 'worst ever' Omega for which no bit can be determined, even with the help of the most powerful formal axiomatic system used by mathematicians, known as Zermelo–Fraenkel set theory.

Each Omega depends on the choice of the computing machine, so there is not just one Omega (as there is only one π), but a class of Omegas. Are there random real numbers other than Omegas that are computably enumerable? This question originates in a 215-page manuscript written by Solovay in 1975 (unpublished manuscript, IBM T. J. Watson Research Center, New York). For years, people working in complexity theory felt that the answer was positive. Solovay imagined a new class of Omega-like numbers that would be distinct from Omegas and would separate

them from the other computably enumerable random real numbers (Fig. 1). The intention was to make an Omega-like number share the paradoxical status of an Omega, but not all the properties of a true Omega; Omega-like numbers would then outnumber the Omegas. An Omega-like real number behaves like an oracle: its huge amount of information can be used to compute close approximations for every computably enumerable real number.

In 1998 a first unexpected result was proved: every Omega-like real number is an Omega⁷. The existence of a computably enumerable random real number that is not an Omega became less plausible, but was not ruled out. The last step has been brilliantly accomplished by another Berkeley mathematician, Theodore Slaman², who has proved that every computably enumerable random real number is Omega-like, and hence an Omega.

This work reinforces the message of algorithmic information theory that randomness is as fundamental and as pervasive in pure mathematics as it is in theoretical physics. In our opinion it also gives further support to 'experimental mathematics', and to the 'quasi-empirical' view of mathematics which says that although mathematics and physics are different, it is more a matter of degree than black and white^{4,5,8}. Physicists are used to working with assumptions that explain a lot of data, but that can be contradicted by subsequent experiments. But mathematicians don't like having to backpedal. Even after Gödel and Turing showed that Hilbert's dream didn't work, in practice most mathematicians carried on as before, in Hilbert's spirit. But now, finally, the computer is changing the way we do things. It is easy to run a mathematical experiment on a computer, but you can't always find a proof to explain the results. So in order to cope, mathematicians are sometimes forced to proceed in a more pragmatic manner, like physicists. The Omega results provide a theoretical underpinning for this revolution. □

C. S. Calude is in the Department of Computer Science, Auckland University, Private Bag 92019, Auckland, New Zealand.

e-mail: cristian@cs.auckland.ac.nz

G. J. Chaitin is at the IBM T. J. Watson Research Center, PO Box 704, Yorktown Heights, New York 10598, USA.

e-mail: chaitin@watson.ibm.com

1. Chaitin, G. J. *J. Assoc. Comp. Mach.* **22**, 329–340 (1975).
2. Slaman, T. A. <http://math.berkeley.edu/~slaman/papers/random.pdf>
3. Solovay, R. M. *Centre for Discrete Mathematics and Theoretical Computer Science Research Report 104* (Auckland Univ., New Zealand, 1999); <http://www.cs.auckland.ac.nz/staff/cgi-bin/mjd/secondcgi.pl>
4. Chaitin, G. J. *The Limits of Mathematics* (Springer, Singapore, 1998).
5. Chaitin, G. J. *The Unknowable* (Springer, Singapore, 1999).
6. Stewart, I. *Nature* **232**, 115–116 (1988).
7. Calude, C. S. et al. in *Proc. 15th Ann. Symp. Theor. Aspects Comp. Sci.* Vol. 1373, 596–606 (Springer, Berlin, 1998).
8. Tymoczko, T. *New Directions in the Philosophy of Mathematics* (Princeton Univ. Press, 1998).

Daedalus

Fiery compositions

Combustion is seldom a steady process. Most flames oscillate in the sub-audio region; singing flames and some pyrotechnic compositions whistle at much higher frequencies. Daedalus now hopes to tame such oscillations for entertainment purposes. Combustion energies are so high that even a modest rate of burning could generate intense sound.

DREADCO's chemists are now exploring this idea. Their pilot project is simply a firework consisting of alternate layers of fast-burning and slow-burning composition. As the combustion zone moves down through the layers, the gas-pressure generated by burning will be modulated, and the firework will howl out a predictable tone. But how to record an audio signal on such a firework? An electrolytic process seems hopeful. A stick of pyrotechnic composition, dampened to make it conducting, will be passed between a pair of electrodes carrying the analogue audio signal to be recorded. With suitably ingenious chemistry, the density of ionic deposition at each point will control the rate or gas-output of burning at that point. Even with a sonic efficiency of a few per cent (about as good as most loudspeakers), such a 'sound-stick' could deafen its audience with hundreds of watts of sound. The technology would be ideal for public musical performances, and also as a fire alarm. A suitably encoded sound-stick, lit by the blaze, would bellow out its location and appeal urgently for help.

Seeking a milder and cheaper version of the idea, Daedalus recalls the old trick of painting a track on a piece of paper with potassium nitrate solution. When later ignited, the track is revealed by a smouldering glow which travels along it. He proposes to print such a track on paper in directly digital form, as minute dots of oxidant scaled in binary intensity. When lit, the combustion zone would speak the digital signal as it traversed its 'sound track'. Audio burn papers, each with many sound tracks, could be printed in large numbers cheaply and easily.

This elegant technology will combine utter simplicity with a useful sound output. Even a few watts of glow will make a respectable noise. Each track can be played once only, of course, but a paper could carry many tracks, and a music-lover could carry many papers. His inventory of pestilent personal electronic gadgets — mobile phone, pager, calculator and so on — would at least be reduced by one.

David Jones

Princess of palaeontology

Mary Anning came from humble origins to find fame as a fossil hunter on England's south coast. Others took credit for her discoveries but it was said of her that she understood more of the science than "anyone else in the kingdom".

Crispin Tickell

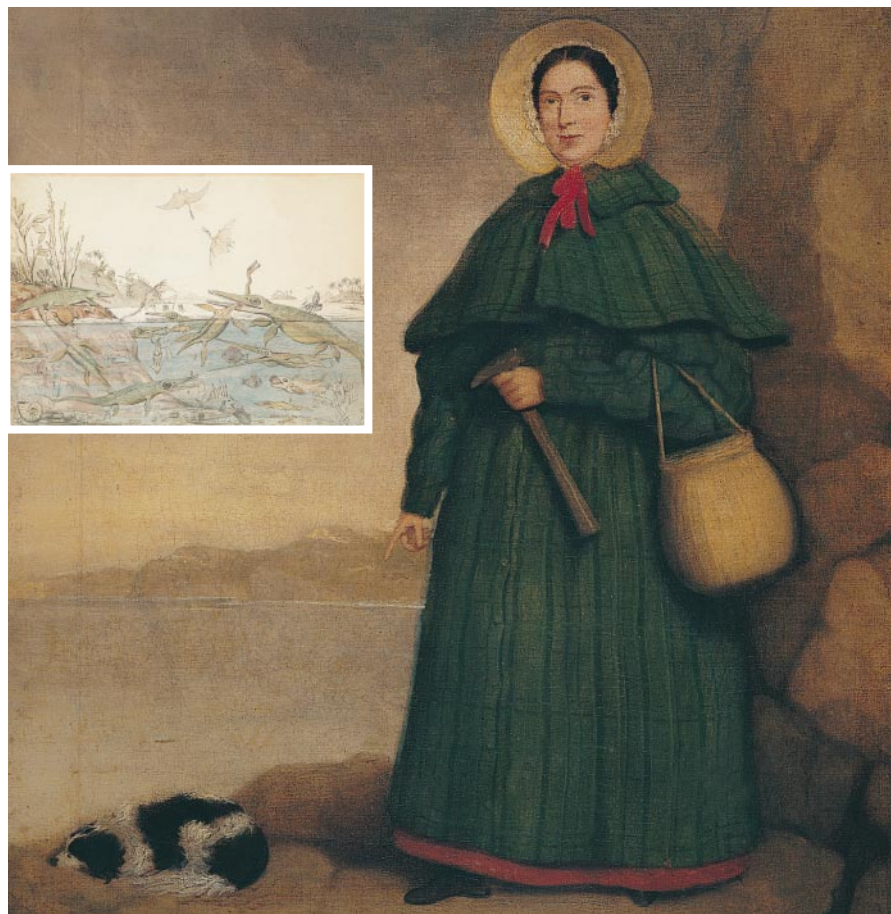
This year is the bicentenary of the birth of Mary Anning of Lyme Regis. She has become famous first as the discoverer — and saleswoman — of the exotic fossil creatures of the Lower Jurassic; and secondly as the prettified heroine of many children's books, sometimes identified with the "who sells sea-shells on the sea shore". A recent symposium in her honour at Lyme Regis brought together palaeontologists, historians and sociologists of science.

Anning was the daughter of a carpenter or cabinet-maker. Although she wrote that she was illiterate, this was far from so. Her letters were clearly written, and she is said to have taught herself French in order to read Georges Cuvier. Her father collected and sold 'curiosities' from the Dorset cliffs and beaches, and when he died in 1810 he left her, his son Joseph and his widow Molly the rudiments of a business. It seems to have been Joseph who uncovered the head of the famous ichthyosaur, which he and Mary excavated and sold.

It requires a major effort of the imagination to think back into the mental world of the 1820s. How did she interpret the world around her, and the place in it of the fossils she found? At one level far from hers, the concept of deep time was just beginning to take hold. In 1795 James Hutton set out the notion of a past and present in which "we find no vestige of a beginning, no prospect of an end". But for most people, including those in rural Dorset, the Bible derived directly from God and was the overriding authority on the history of the Earth (notoriously begun, according to Archbishop Ussher, at 9 a.m. on 26 October 4004 BC). Fossil creatures deep in the cliffs with no apparent living descendants presented an obvious difficulty. It was suggested that they had been entombed by the Flood, or even left around by God as a test of people's faith.

The span of Anning's life, from 1799 to 1847, was an age when geology was seen as the queen of the sciences. Anning herself was described as "the princess of palaeontology" in 1837. After despairing efforts to link the fossil evidence with biblical chronology, most geologists gave up, and the more pious were content to interpret the Bible as metaphor. On one side were the uniformitarians of the school of Hutton, William Smith and Charles Lyell; and on the other the catastrophists, often with a theological agenda, who saw history as a series of relatively sudden events or interventions. By the time Anning died, the uniformitarians were in the ascendancy.

Most of the leading geologists of the day —



Anning: her talent for finding and evaluating fossils won her tributes from eminent geologists of the day.

William Buckland, Henry De La Beche, Rodrick Murchison, Adam Sedgwick, Louis Agassiz and others — visited Lyme Regis, and knew Anning. Some came as customers for her fossils, and made expeditions with her. She must have been aware of current controversies and, as Lady Silvester wrote in 1824, "she has arrived at that degree of knowledge as to be in the habit of writing and talking with professors and other clever men... and they all acknowledge that she understands more of the science than anyone else in the kingdom".

She certainly recognized the enormous antiquity of the fossils, and doubted whether they had living counterparts. Theory she left to others. Cautious scepticism was one of her abiding characteristics.

Her strength was practical. She had an extraordinary talent for finding, assembling and evaluating fossils. The list is impressive: four ichthyosaurs, two plesiosaurs, a pterodactyl, the fossil fish *Squaloraja*, plus ammonites, sea urchins and other marine fossils. It was she who recognized coprolites as fossil faeces, and who, by comparing existing

sea hares with fossil belemnites, worked out the association of the belemnite with its ink bag. By grinding up the sepia from a fossilized belemnite ink bag, she made ink for her friends to use in sketching fossils.

Anning suffered from every social handicap. She was poor for all but her last few years, when friends secured a modest pension for her. As a woman she had no prospect of professional recognition. Almost all her discoveries bear the names of those who bought specimens from her. As late as 1930 a local historian described her as no more than a "handmaid of scientific men".

But that was not the view of the scientific men. On her death De La Beche read out an unprecedented tribute to her at the Geological Society. No wonder that she has become something of an icon. Looking at the traces she left, she emerges as a tough, independent-minded person of skill and intelligence who surmounted her circumstances to help lay the foundations for a new science of the Earth. □

Crispin Tickell is at Abington Old Barn, Abington, Cirencester, Gloucestershire GL7 5NU, UK.

Magnet levitation at your fingertips

The stable levitation of magnets is forbidden by Earnshaw's theorem, which states that no stationary object made of magnets in a fixed configuration can be held in stable equilibrium by any combination of static magnetic or gravitational forces¹⁻³. Earnshaw's theorem can be viewed as a consequence of the Maxwell equations, which do not allow the magnitude of a magnetic field in a free space to possess a maximum, as required for stable equilibrium. Diamagnets (which respond to magnetic fields with mild repulsion) are known to flout the theorem, as their negative susceptibility results in the requirement of a minimum rather than a maximum in the field's magnitude²⁻⁴. Nevertheless, levitation of a magnet without using superconductors is widely thought to be impossible. We find that the stable levitation of a magnet can be achieved using the feeble diamagnetism of materials that are normally perceived as being non-magnetic, so that even human fingers can keep a magnet hovering in mid-air without touching it.

Stable levitation has been demonstrated for diamagnetic objects such as superconducting pellets and live creatures^{2,5-7}. Strong diamagnetism of superconductors allows the situation to be reversed, so that a magnet can be levitated above a superconductor⁸. Paramagnetic objects can also be levitated if placed in a stronger paramagnetic medium, such as ferrofluid or oxygen, which makes them effectively diamagnetic⁹.

We set out to lift a magnet by applying a magnetic field and then stabilizing the intrinsically unstable equilibrium with repulsive forces from a nearby diamagnetic material. We found that, surprisingly, the forces created by almost non-magnetic materials (susceptibility χ of about 10^{-5}) are sufficient to stabilize levitation over distances as large as several millimetres under Earth gravity conditions, even though they decay rapidly with distance as $1/x^5$ (Fig. 1).

For stable levitation, an equilibrium requires that the magnetic force $MB'(z)$ compensates the gravitational force mg , where M is the magnetic moment and $B(z)$ and $B'(z)$ are the magnetic field on the axis and its derivative, respectively. For the equilibrium to be stable, it must be in a region where the total energy of the magnet $U = -MB(r) + mgz + U_{\text{dia}}$ has a minimum ($\Delta U > 0$), where U_{dia} is the energy of diamagnetic interaction with the cylinder. Close to the equilibrium position at the field axis^{3,10},

$$U \approx U_0 + [mg - MB'(z)]z + K_v z^2 + K_h r^2 + \dots \quad (1)$$

where $K_v(z) \equiv -MB''(z)/2$ and

$$K_h(z) \equiv -M[B'(z)^2 - 2B(z)B''(z)]/8B(z).$$

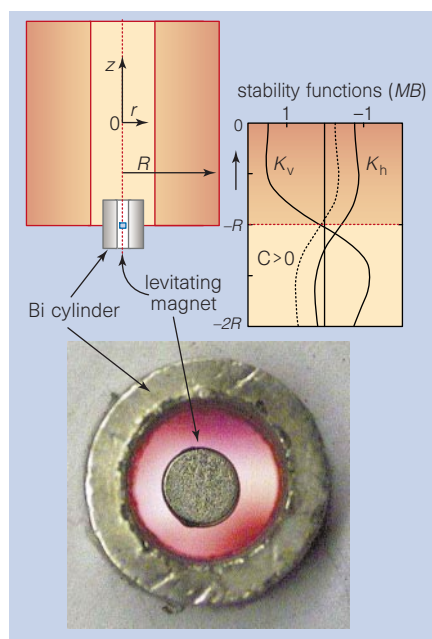


Figure 1 A NdFeB magnet (an alloy of neodymium, iron and boron; 4 mm high and 4 mm in diameter) levitating at the axis of a vertical solenoid of radius $R \approx 10$ cm and length $\approx 2R$ in a magnetic field of 100 gauss. The levitation is stabilized by a bismuth cylinder ($\chi = -1.5 \times 10^{-4}$) with inner diameter $D \approx 8$ mm. The photograph shows the top view of the levitating magnet. The right-hand plot shows the stability functions K_v and K_h , calculated for a solenoid with a height of twice its radius (solid curves). Diamagnetic interaction C shifts the horizontal stability function K_h to the left (dashed curve) and a small region of positive ΔU emerges above the point where $K_v = 0$.

The presence of a diamagnetic cylinder results in the last term in equation (1) and, for the geometry of Fig. 1, we find that $C = 45\mu_0\chi M^2/16D^5$, where μ_0 is the permeability of free space. If there is no diamagnet ($C = 0$), the stability can never be reached (at no point are K_v and K_h both positive; Fig. 1). The diamagnetic interaction allows the energy U to have a minimum ($K_v > 0$ and $K_h + C > 0$) which emerges for $C > MB'(z)^2/8B(z)$ just above the point of a maximum field gradient ($B''(z) = 0$). It is counterintuitive that levitation is easiest in the most inhomogeneous field region, rather than in the centre of a solenoid where the field is almost uniform.

It is instructive to introduce a characteristic scale L on which the field changes: $B' = B/L$. At the optimum levitation point ($B''(z) = 0$), L varies between R and $1.2R$ for long and short solenoids, respectively. If we approximate our levitating magnet by a sphere of diameter d with a remnant field B_r , then $M = (\pi/4\mu_0)B_r d^3$, and the requirement for levitation becomes

$$A(|\chi|LB_r^2 d^3/\mu_0 \rho g)^{1/5} > D > d \quad (2)$$

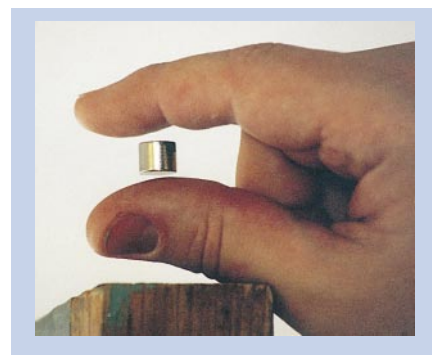


Figure 2 Levitation at your fingertips. A strong NdFeB magnet (1.4 tesla) levitates 2.5 metres below a powerful superconducting magnet. The field at the levitation point is about 500 Gauss.

where ρ is the density of the magnet's material and $A \approx 1.92$. Calculation shows that a magnet several millimetres in size with remnant magnetization of about 1 tesla (NdFeB) can be levitated with a clearance gap, $D - d$, of several millimetres using a 10-cm solenoid and strongly diamagnetic Bi or graphite, in agreement with our experiment.

Equation (2) depends on the product of χ and L , which means that by increasing L (scaling up the magnet's size) we can achieve the same D as above using ordinary materials (such as plastic or wood, with $\chi \approx -10^{-5}$). To illustrate this point, we show another example of a levitating magnet in Fig. 2 in which human fingers ($\chi \approx -10^{-5}$) are used as diamagnetic stabilizers. Here we use an alternative geometry¹¹ in which L is easier to scale up because it is determined not only by the magnet size, but also by its strength. The levitating magnet is placed below a solenoid in the region where the equilibrium is stable horizontally ($K_h > 0$) but not vertically ($K_v < 0$) (Fig. 1). Vertical stability is achieved by means of two horizontal diamagnetic plates (or by the fingertips).

In this geometry, the positive constant $C = 6\mu_0\chi M^2/\pi D^5$ counters K_v and the levitation condition is similar to equation (2), except that now D denotes the separation between the plates, $A \approx 1.02$ and $L \approx 4B'/B''$ is approximately the distance from the centre of a solenoid to a levitating magnet. The larger the distance, the easier it is to stabilize levitation by diamagnetic repulsion. L is limited by the requirement on the field gradient, $B'(z) = mg/M$. To reach such a large L , as in Fig. 2, we used an 11-tesla superconducting solenoid a metre in diameter. If stronger diamagnets are used (such as graphite or bismuth), this type of levitation can also be achieved with small permanent magnets, making miniature hand-held

devices accessible to everyone (M. D. S., unpublished data). These could replace the existing servo levitation devices for some applications.

A. K. Geim*, M. D. Simon†, M. I. Boamfa*, L. O. Heflinger†

*High Field Magnet Laboratory,

University of Nijmegen,

6525 ED Nijmegen, The Netherlands

e-mail: geim@sci.kun.nl

†Department of Physics and Astronomy,

University of California at Los Angeles,

Los Angeles, California 90095, USA

1. Earnshaw, S. *Trans. Camb. Phil. Soc.* **7**, 97–112 (1842).

2. Brandt, E. H. *Science* **243**, 349–355 (1989).

3. Berry, M. V. & Geim, A. K. *Eur. J. Phys.* **18**, 307–313 (1997).

4. Thomson, W. (Lord Kelvin) *Reprints of Papers on Electrostatics and Magnetism* (Macmillan, London, 1872).

5. Braunbeck, W. *Z. Phys.* **112**, 753–763 (1939).

6. Beaugnon, E. & Tournier, R. *Nature* **349**, 470 (1991).

7. Geim, A. *Phys. Today* **51**, 36–39 (September 1998).

8. Arkadiev, A. *Nature* **160**, 330 (1947).

9. Ikezoe, Y. *et al. Nature* **393**, 749–750 (1998).

10. Simon, M. D. *et al. Am. J. Phys.* **65**, 286–292 (1997).

11. Boerdijk, A. H. *Philips Tech. Rev.* **18**, 125–127 (1956).

Extraocular magnetic compass in newts

Geomagnetic orientation is widespread among organisms, but the mechanism(s) of magnetoreception has not been identified convincingly in any animal¹. In agreement with biophysical models proposing that the geomagnetic field interacts with photoreceptors^{2–4}, changes in the wavelength of light have been shown to influence magnetic compass orientation in an amphibian, an insect and several species of birds (reviewed in ref. 5). We find that light-dependent magnetic orientation in the eastern red-spotted newt, *Notophthalmus viridescens*, is mediated by extraocular photoreceptors, probably located in the pineal complex or deeper in the brain (perhaps the hypothalamus).

Experiments investigating shoreward magnetic compass orientation have demonstrated that the newt's perception of the direction of the magnetic field is rotated 90° under long-wavelength (greater than 500 nm) light^{5,6}. We recently trained newts under natural skylight to aim for the shore by placing them for 12–16 hours in water-filled tanks with an artificial shore at one end^{5,7}. The magnetic orientation of individual newts was then tested in a circular, visually symmetrical indoor arena under depolarized light. Under full-spectrum light (from a xenon arc source), they exhibited bimodal magnetic orientation parallel to the shoreward axis in the training tank (Fig. 1a,d). In contrast, under long-wavelength light, they orientated themselves perpendicular to the shoreward direction (Fig. 1b,e).

To demonstrate that the 90° shift in orientation under long-wavelength light was

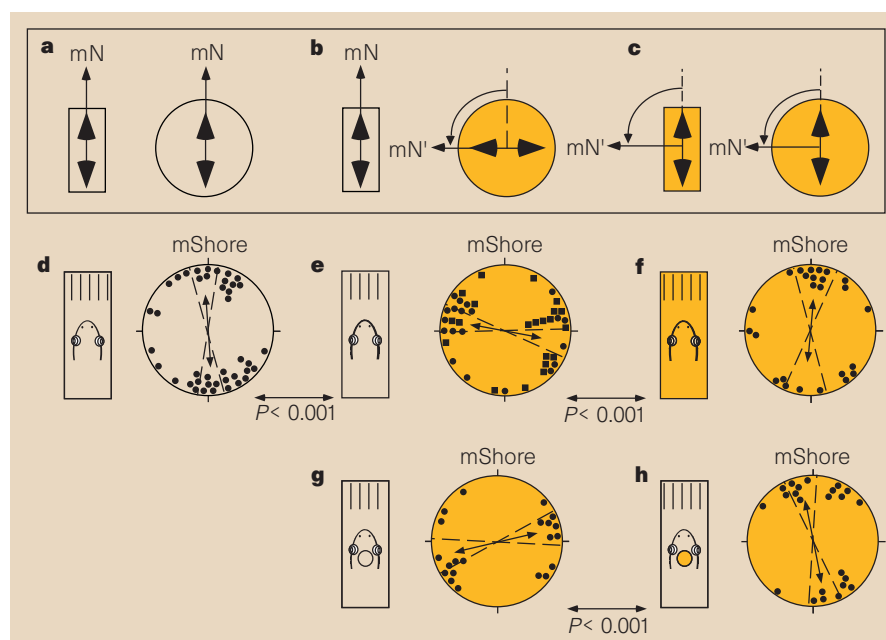


Figure 1 Effects of long-wavelength light and head caps on bimodal magnetic orientation in newts. **a–c**, Predicted orientation of newts (double-headed arrow) and their perception of the direction of the magnetic field (single-headed arrow)^{5,6}. Training tanks have the shore towards magnetic north (mN); circular test arenas show the predicted response of the newts under either full-spectrum (beige) or long-wavelength light (yellow). **a**, Full-spectrum training and testing: newts should perceive the shore to be towards magnetic north and exhibit bimodal magnetic orientation along the shoreward axis. **b**, Full-spectrum training, long-wavelength testing: newts' perception of magnetic north in testing, and their orientation in the test arena, should be rotated 90° (mN') from magnetic north during training. **c**, Long-wavelength training and testing: newts' perception of the magnetic field should be rotated 90° relative to the actual field during training and testing. Their perception of the magnetic field in the arena would be the same as in the outdoor tank. **d–h**, Results. Data points show the magnetic bearing of a newt tested in one of four symmetrical alignments of an Earth-strength magnetic field (magnetic north is geographic north (gN), east (gE), west (gW) or south (gS)). The magnetic field was altered by two orthogonally orientated, double-wrapped, Ruben's coils around the test arena⁷. Data are plotted with respect to the magnetic direction of shore (mShore) in the training tank (shore direction, 360°). Double-headed arrows indicate mean axis of orientation with the mean axis length, r , proportional to the strength of orientation (diameter of the circle corresponding to $r=1$). Dashed lines indicate 95% confidence intervals for the mean axis. Distributions are significant at $P<0.05$ or less by the Rayleigh test and P -values between circle plots indicate significant differences between distributions (Watson U^2 test). **d**, Newts trained under natural light and tested under full-spectrum light orientated along the shoreward axis⁵. **e**, Newts trained under natural light and tested under long-wavelength light orientated 90° from the shoreward direction (filled circles; tested under broadband long-wavelength (≥ 500 nm) light; filled squares, tested under a 550-nm light, 40 nm bandwidth, $12.5 \pm 0.1 \log \text{Quanta cm}^{-2} \text{s}^{-1}$; ref. 5). **f**, Newts trained and tested under long-wavelength light orientated along the shoreward axis⁵. **g**, After training under natural light, clear-capped newts tested under long-wavelength light orientated $\sim 90^\circ$ from the shoreward direction. **h**, After training under natural light, newts with long-wavelength-transmitting caps orientated along the shoreward axis under long-wavelength light.

due to a direct effect of light on the newts' perception of the magnetic field, we trained newts under long-wavelength light by covering the training tank with a long-wavelength-transmitting gel filter (two layers of Lee #101)⁵. Under long-wavelength light, these newts orientated themselves parallel to the shoreward axis, indicating that they had learned the direction of the shore with respect to the rotated magnetic information under long-wavelength light (Fig. 1c,f).

As well as ocular photoreceptors, newts have extraocular photoreceptors in the pineal complex⁸ and possibly the hypothalamus⁹. To determine which photoreceptors are involved in the magnetic compass response, we manipulated the wavelength of light reaching the extraocular photorecep-

tors. Small round 'caps' (5 mm in diameter) were attached to the dorsal surface of the head of each newt using cyanoacrylate glue, and remained in place during both training and testing. Equal numbers of newts were capped with either a clear filter (Lee #130) or a filter that transmitted only long-wavelength light (equivalent to two layers of Lee #101). The caps were positioned to alter the spectral properties of light reaching the pineal and surrounding structures, whereas light reaching the eyes was unaffected. Clear-capped newts were tested to control for any nonspecific effects of the caps on the newts' orientation behaviour.

All newts were trained outdoors under natural skylight and tested for magnetic orientation in the testing arena under long-

Table 1 Magnetic orientation of newts in the 'cap' experiments

Shore direction	Cap type	Magnetic axis (°)	<i>r</i>	<i>n</i>	<i>P_R</i>	<i>U</i> ²	<i>P_W</i>
West (shore, 270°)	clear	172–352	0.49	11	<0.1	0.346	<0.001
	yellow	77–257	0.54	11	<0.05		
North (shore, 360°)	clear	62–242	0.76	10	<0.001	0.472	<0.001
	yellow	162–342	0.61	12	<0.01		
Pooled data (shore, 360°)	clear	77–257	0.58	21	<0.001	0.753	<0.001
	yellow	169–349	0.57	23	<0.001		

Capped newts were trained to exhibit shoreward magnetic compass orientation in an outdoor tank with the shore direction to either the west or the north, and tested for magnetic orientation as in previous experiments^{5,7} under long-wavelength (≥ 500 nm) light. Magnetic bearings were doubled before statistical analysis. For the pooled distributions (Fig. 1g,h), magnetic bearings from newts trained to the west shore were rotated by $+90^\circ$ so that the shoreward direction for both groups was at 360° . Magnetic axis is the mean for the bimodal distribution; *r*, mean vector length; *n*, sample size; *P_R*, probability by the Rayleigh test; *U*², Watson *U*² test for differences between two distributions; *P_W*, probability by the Watson *U*² test. For statistical analyses of the data shown in Fig.1d–f, see ref. 5.

wavelength light. Newts with clear caps were predicted to orientate along an axis perpendicular to the shoreward direction (like newts trained under full-spectrum light and tested under long-wavelength light; Fig. 1e). In contrast, the predicted orientation of newts with 'long-wavelength' caps depended on whether the caps altered the wavelength of light reaching the photoreceptors involved in the magnetic compass. If so, these photoreceptors would have been exposed to long-wavelength light in both training and testing, and the newts should have orientated along the shoreward axis (as in Fig. 1c, f). If not, the photoreceptors would have been exposed to full-spectrum light during training and long-wavelength light during testing. In this case, the newts should have orientated perpendicular to the shore (as in Fig. 1b,e).

Newts with clear caps orientated perpendicular to the shoreward magnetic axis under long-wavelength light, indicating that the caps did not alter the orientation response (Table 1; Fig. 1g). In contrast, newts with long-wavelength-transmitting caps exhibited bimodal magnetic orientation parallel to the shoreward axis in the training tank (Table 1; Fig. 1h). Covering the dorsal surface of the newt's head with a long-wavelength-transmitting filter therefore mimicked the effect of long-wavelength training on the newt's shoreward compass response. Thus, extraocular photoreceptors are involved in the newt's light-dependent magnetic compass.

The effects of light on magnetic orientation in birds are different from those in newts (birds have been disorientated under wavelengths beyond 590 nm)². Results from physiological studies indicate that both the avian visual system and the photosensitive pineal are sensitive to magnetic fields⁵, but behavioural experiments on birds from which the pineal has been removed suggest that it is not required for avian magnetic compass orientation^{10,11}. It would not be surprising, therefore, if the photoreceptors underlying such orientation are in different locations in salamanders and birds (extraocular and ocular, respectively). Alternatively, our findings and the results of experiments on pinealectomized birds^{10,11}

could both be explained by the involvement of deep brain photoreceptors in a light-dependent magnetic compass.

M. E. Deutschlander*, S. C. Borland, J. B. Phillips

The Center for the Integrative Study of Animal Behavior, Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

*Present address: Department of Biology, University of Victoria, P.O. Box 3020, Victoria, British Columbia V8W 3N5, Canada e-mail: mdeutsch@uvic.ca

1. Wiltshcko, R. & Wiltshcko, W. *Magnetic Orientation in Animals* (Springer, Berlin, 1995).
2. Schulten, K. in *Advances in Solid State Physics* (ed. Treusch, J.) 61–83 (Vieweg, Braunschweig, 1982).
3. Leask, M. J. M. *Nature* **267**, 144–145 (1977).
4. Edmonds, D. T. *Proc. R. Soc. Lond. B* **263**, 295–298 (1996).
5. Deutschlander, M. E., Phillips, J. B. & Borland, S. C. *J. Exp. Biol.* **202**, 891–908 (1999).
6. Phillips, J. B. & Borland, S. C. *Nature* **359**, 142–144 (1992).
7. Deutschlander, M. E., Phillips, J. B. & Borland, S. C. *Copeia* (in the press).
8. Adler, K. *Photochem. Photobiol.* **23**, 275–298 (1976).
9. Foster, R. G., Grace, M. S., Provencio, I., DeGrip, W. J. & Garcia-Fernandez, J. M. *Neurosci. Biobehav. Rev.* **18**, 541–546 (1994).
10. Schneider, T., Thalau, H.-P., Semm, P. & Wiltshcko, W. *J. Exp. Biol.* **194**, 255–262 (1994).
11. Maffei, L., Meschini, E. & Papi, F. *Z. Tierpsychol.* **62**, 151–156 (1983).

Dating the origin of HIV-1 subtypes

The moment in history when subtypes of the human immunodeficiency virus HIV-1 became distinguishable is hotly debated^{1,2}. Zhu *et al.*³ have provided a unique HIV-1 sequence from 1959, known as ZR59. Based on the position of ZR59 in phylogenetic trees and its distance from the common node of the viral B/D/F-subtype lineages, they suggest that subtypes B and D have evolved from a single introduction "not long before 1959", and that HIV-1 group M viruses probably shared a common ancestor "in the 1940s or the early 1950s". Here we caution against such precise dating of these evolutionary events.

Evolutionary rates among viruses are unlikely to be equal and constant over time, as the sequences presented by Zhu *et al.*

testify, with subtype B sequences from 1983–94 being significantly closer to the B/D/F node than are subtype D sequences sampled in the same period. We have retrieved 95 HIV-1 strains (75 B, 12 D and 8 F) from the GenBank database, for which our only selection criteria were that sequences covered the same genetic regions as ZR59_{abc} and that sampling years were reported (1983–97). A linear regression analysis revealed that there was a significant ($P < 0.001$) but weak ($r^2 = 0.07$) positive correlation between sampling years and sequence distances to the B/D/F node, with a slope (divergence rate) of 0.0009 or 0.0008 (including or excluding ZR59, respectively).

Although a molecular clock is operational during HIV-1 evolution in an AIDS pandemic⁴, it is of limited value for dating the B/D/F node or the ancestor of group M viruses⁵: when we estimated the year the B/D/F lineages diverged, despite the 95% confidence interval of more than 100 years, we arrived at the year 1913 by assuming that each sequence was statistically independent and that HIV-1 evolution was not influenced by population bottlenecks. As there is no reliable way to control for the validity of these two assumptions and to narrow the confidence interval, such calculations can be dismissed as uninformative.

When the limitations of the molecular clock are ignored and no confidence intervals for node dates are derived, all we have to work with is the genetic distance of the ZR59 sequence to the nodes. The mean sampling year of the contemporary sequences was 1990 ± 5 (s.d.) and their mean distance to the B/D/F common node was 0.071 ± 0.017 , which is only 2–3 times higher than we and Zhu *et al.*³ estimate for the ZR59 sequence. Even if the line of reasoning pursued by Zhu *et al.* is accepted, the time between the B/D/F node and 1959 should be equal to or half that passed between 1959 and 1990, leading to a date for the B/D/F node of either 1928 or 1944.

For the group M node, a date decades before that is possible, but it is hard to date as its position depends on the phylogenetic method used. In our analysis, branching patterns of ZR59 and the position of the group M node (Fig. 1a) are similar to those of Zhu *et al.*, but other methods yield a completely different position for the M node⁵.

We uncovered several contemporary HIV-1 sequences that are equidistant or closer than the ZR59 sequence to the nodes of group M (Fig. 1b) and B/D/F lineages (Fig. 1c). Their existence makes dating of the nodes shortly before 1959 more cumbersome. Variation among HIV-1 sequences is increasing over time yet, in any year, sequences can be found that are close to the ancestor⁴. The Poisson model of the molecular clock implies variation in evolution rates among lineages^{4,6}. Together with massive

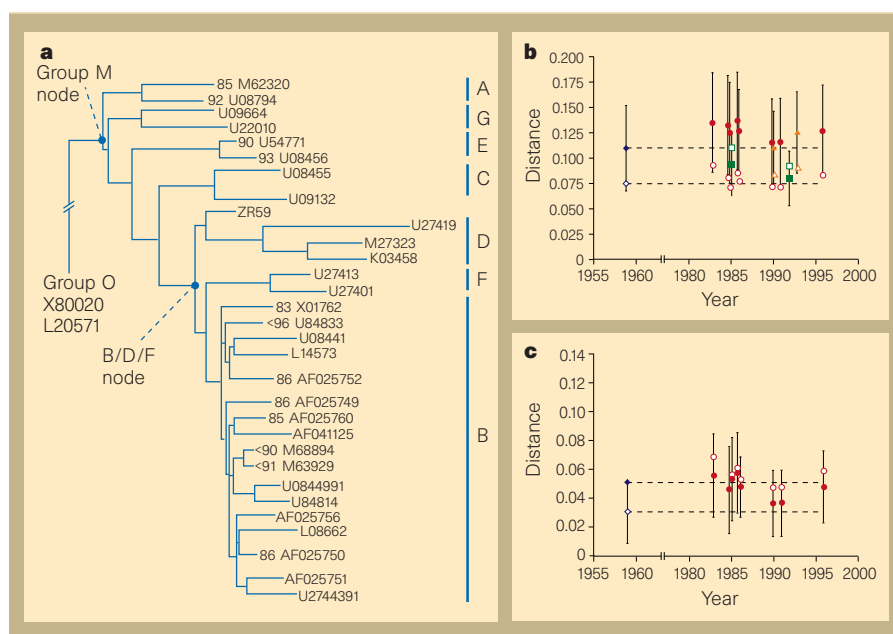


Figure 1 Analysis of HIV-1 sequences. **a**, Phylogenetic analysis of ZR59 and virus strains representing HIV-1 group M subtypes A–G and group O using the maximum-likelihood method. Sequences used by Zhu *et al.*³ are included as subtype references, as well as sequences closest to the B/D/F node. Trees were produced with the DNAML program from the PHYLIP package using empirically found base frequencies and transition/transversion ratios, considering different rates of evolution at different positions, and using group O viruses as an outgroup. Sequences are labelled by GenBank accession number and HIV-1 subtype; sampling years are shown for the sequences used in **b** and **c**. The neighbour-joining method using gamma distances for the Jukes-Cantor and Kimura two-parameter models has produced trees with similar topology (results not shown). **b**, **c**, Genetic distances of HIV-1 sequences to the group M (**b**) and the B/D/F (**c**) common nodes in relation to sequence sampling years, according to the maximum-likelihood (filled symbols; vertical bars show distance confidence limits) and neighbour-joining (open symbols) methods. Sequences closest to the nodes are shown. Circles, subtype B sequences; squares, subtype A sequences; triangles, subtype E sequences; diamonds and dashed lines, ZR59 sequence.

population bottlenecks in the history of HIV, this allows for the circulation of viruses close to the B/D/F node 30 or more years before, as well as after, 1959. Neither can the presence in 1959 of HIV-1 strains belonging to modern HIV-1 subtypes be excluded.

The clock-like nature of HIV-1 evolution is evident only for intrasubtype synonymous evolution during expansion of the AIDS epidemic^{4,5}. The nature of this clock is difficult to reconcile with the point of view that an HIV-1 sequence that is relatively close to a reconstructed ancestor inevitably circulated only for a short period after the ancestor. As evolutionary rates among HIV-1 lineages are highly variable, and its history is strongly influenced by population bottlenecks and founder effects, a precise dating based on a single sequence of the B/D/F ancestor to a few years before 1959, and of the group M ancestor to the 1940s or 1950s, is probably premature.

Jaap Goudsmit, Vladimir V. Lukashov

Department of Human Retrovirology,
Academic Medical Center,
University of Amsterdam, Meibergdreef 15,
1105 AZ Amsterdam, The Netherlands

- Myers, G. *AIDS Res. Hum. Retroviruses* **10**, 1317–1324 (1994).
- Goudsmit, J. *Viral Sex: The Nature of AIDS* (Oxford Univ. Press, 1997).

- Zhu, T. *et al. Nature* **391**, 594–597 (1998).
- Lukashov, V. V. & Goudsmit, J. *J. Virol.* **71**, 6332–6338 (1997).
- Korber, B., Theiler, J. & Wolinsky, S. *Science* **280**, 1868–1871 (1998).
- Lukashov, V. V., Kuiken, C. L. & Goudsmit, J. *J. Virol.* **69**, 6911–6916 (1995).

Korber et al. reply — We remain confident that our study¹ shows that an M group HIV-1 was circulating in humans in 1959 and that subtypes B, D and F arose after that. Our estimates were based on genetic distances from ZR59 to the common node of B, D and F, and to the centre of the ‘starburst’ phylogeny for all HIV-1 isolates¹. We cautioned, however, that genetic distances cannot be precisely translated into years for a single virus because of the variation in the rate of evolution among different individuals and different lineages.

We question several points made by Goudsmit and Lukashov. First, there appears to be some bias in their sample selection: for example, they cite an average genetic distance to the common node of subtypes B, D and F of 0.071 among 95 HIV-1 isolates of these three subtypes, a value that is not consistent with those plotted in Fig. 1c.

Second, their argument is based on the closer proximity, compared with ZR59, of

some contemporary HIV-1 sequences to the common node of all M-group viruses. However, the position of this node, where the O group branch joins the M group, is unreliable for phylogenetic analysis of HIV-1. In their tree, it comes into the M group on a branch leading to subtype A, but this placement might vary depending on the phylogenetic reconstruction method, the region of the genome under study, and the particular O group sequences selected. For example, in an analysis of the sequences of 113 full-length gp160 glycoproteins, the simian immunodeficiency virus CPZ sequence enters the tree on a branch to subtype C (ref. 2). This ambiguity in rooting M group viruses was why we did not estimate the timing of the origin of the M group viruses more precisely, and it may undermine the data plotted in Fig. 1b.

Third, the error bars shown in Fig. 1b,c may be overestimates. Although PHYLIP calculates the error on individual branches, it is not trivial to combine these errors when adding branch lengths. Our errors were asymmetrical and considerably less than the confidence interval of 100 years suggested by Goudsmit and Lukashov. Their estimated errors allow no more than 95% confidence that M group viruses originated in humans before 2013.

We agree¹ that there are limitations in assuming a molecular clock to estimate the timing of a particular node in a phylogenetic tree (reviewed in ref. 2), but we still contend that our conclusions are plausible and supported by the data. The branch length of ZR59 to the ancestral node of subtypes B, D and F was quite short, and comparable to the distances between viral sequences from patients with related infections. Phylogenetic tree algorithms are less subject to bias when estimating short branch lengths.

From the short branch length of ZR59 to the ancestral node of subtypes B, D and F, we inferred that the ancestral virus for subtypes B, D and F existed only a few years before 1959, and probably not more than 15 years earlier. This does not assume that M group evolutionary rates are the same, only that they are roughly of the same order as the evolutionary rates we are seeing in the HIV-1 epidemic, which is a reasonable and parsimonious assumption.

Bette T. Korber*, Paul M. Sharp†, David D. Ho‡

*Los Alamos National Laboratory,
Los Alamos, New Mexico 87545, USA

†Department of Genetics,
University of Nottingham, Queens Medical Centre,
Nottingham NG7 2UH, UK

‡Aaron Diamond AIDS Research Center,
The Rockefeller University, New York,
New York 10016, USA

- Zhu, T. *et al. Nature* **391**, 594–597 (1998).
- Korber, B., Theiler, J. & Wolinsky, S. *Science* **280**, 1868–871 (1998).

Under the covers or under a lens

Designing babies

by Roger Gosden

W. H. Freeman: 1999. 260 pp. \$24.95

Carl Djerassi

Both the author of this book and readers ought to know my personal biases from the outset. As a chemist, deeply involved and interested in human conception over four decades, I have become so intrigued during the past few years by the cultural and operational effects of assisted reproductive technologies, including IVF techniques such as intracytoplasmic sperm injection (ICSI), and the separation of sex (under the covers) and fertilization (under the microscope), that I even wrote a novel and a play about the subject. So I opened Gosden's book with a tinge of competitive curiosity, even looking for authorial mis-steps or omissions. But I closed it with unambiguous admiration. "Read this book!" is my warm recommendation, because these topics are covered well.

Gosden surveys in a readable, breezy fashion all the high points of the reproductive revolution that started 21 years ago in Britain with the birth of the first IVF baby, Louise Joy Brown. His prose is straightforward enough not to lose general readers but not so simple as to turn off scientifically sophisticated ones. He is remarkably up-to-date, covering research results until late 1998, as can be seen from an admirable bibliography. And he addresses, perhaps a tad too optimistically, all the contentious issues, from sex predetermination to preimplantation analysis and cloning.

His speculations on how to get men pregnant, however, even by equipping them with uteri, is likely fodder for reproductive Ludites who will simply exclaim "I told you so" and not pay attention to the other issues raised here, 98 per cent of which are truly important. I am not sure whether getting men pregnant and converting them into functioning androgenic mothers is currently a high priority for anybody but a tabloid newspaper. But we do need a vigorous debate on the implications of 100 per cent certain sex predetermination, once effective separation of Y and X sperm has become a routine procedure in humans (a topic thoroughly discussed by Gosden). Or on the *fait accompli* of using sperm aspirated after death from a man — even after storing it for months or years — to produce a live child via ICSI.

The best written, and in my opinion also most significant, section is the one dealing with the issue of ageing mothers. In the Western and increasingly geriatric world, first-time motherhood is being postponed. A woman is born with her supplies of immature eggs, 90 per cent of which are gone by the



Prodigy synthesis: reproduction taken to its limit.

time she is 35. As Gosden states succinctly: "Nature has carried out an act of biological sabotage on women" in that the ovary ages faster than any other organ — even though the uterus doesn't. Postmenopausal pregnancies are thus perfectly feasible, provided a source of young, healthy eggs is still available.

The author covers in detail the various options for accomplishing that aim, from surrogate eggs to the preservation of ovarian tissue and its possible subsequent transplantation. But he seems rather dismissive of the already feasible preservation of young mature eggs with concomitant use of IVF. He tends to downplay that approach because of the expense associated with any IVF technology, yet ignores the fact that still-unrealized approaches such as ovarian transplant are likely to be even more costly. He makes a convincing case for why women should be entitled to the choice of later motherhood, even beyond the menopause. The real problem with all of these approaches is, of course, that for a long time to come only the affluent people in the affluent countries will be able to take advantage of such options.

The book is replete with examples of technology being developed for one purpose, but then used by society for very different ones. The direct injection of a single sperm into an egg under the microscope (ICSI), followed by transfer of the two-day-old embryo back into the woman's uterus was developed by Belgian scientists in 1991 to treat male infertility, primarily in men with insufficient sperm. The first baby born of such a technique is now only eight years old, but more than 10,000 such babies have

been born since then and by no means all the fathers have been lacking sperm.

In the end, the reproductive optimists will have to accept Gosden's statement, so typical of many practising specialists in reproductive medicine: "I believe the primary goal of reproductive medicine and biology is to help people bear a child with the best possible start in life — physically and psychologically — and to bring diversity to family life." We all know that life is a terminal disease and that the production of genetically identical offspring is the only indirect way of prolonging it. Gosden makes a good case for how far science, technology and apparently also society are willing to go to satisfy that drive — a subject meriting continuing debate.

Designing Babies may soon require a new edition, given the rapid advances in reproductive medicine, in which case I have two recommendations. The present index is so inadequate that it should be improved or deleted. There are entries for the Nazi party, Shakespeare and Woody Allen, but not for ApoE, BRCA1, PGD or dozens of other highly germane words that are discussed in the text. The glossary is better, but makes no mention of many IVF techniques such as GIFT and SUZI, and makes the error — grating to a chemist — of defining "oestrogen" as "a steroid hormone produced by the ovary". Oestrogen is a generic classification, but there is no such specific hormone (in contrast to oestradiol, oestriol and the like). While one may forgive the popular press for such a misuse, an expert of Gosden's calibre should not perpetuate it. Nor should the author indulge in one-upmanship with buddy references to "Ian", "Steve" and "Arnne". I will offer a good bottle of Californian wine to the first *Nature* reader who can identify all three of Roger's pals without first having read the book. □

Carl Djerassi is in the Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA.

Life as we don't know it

Worlds Without End: The Exploration of Planets Known and Unknown

by John S. Lewis

Perseus: 1998. 236 pp. \$24, £16.50

Joseph A. Burns

The highly controversial claim three years ago that early microbial life once resided in the martian meteorite ALH84001 stunned most scientists with its audacity. But, viewed in another way, this 'evidence' was simply another step in an accelerating continuum of

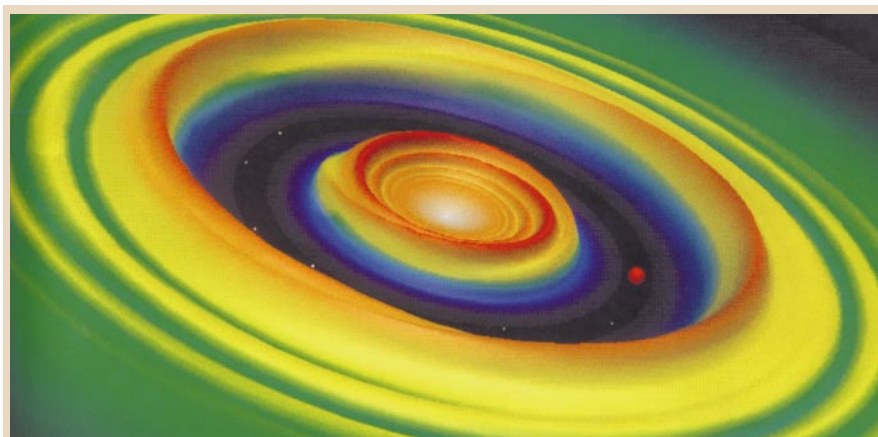
findings that suggest extraterrestrial life might be ubiquitous. After all, space images had been interpreted as indicating that Mars and Europa might be or have been suitable abodes for life. And, starting in late 1995, the first planets around main-sequence stars were discovered; dozens, including members of another planetary system, are now known.

During the past decade, biologists also contributed to the controversy with their remarkable success in constructing the genealogical roots to the family tree of life, which led to the inference that terrestrial life arose rapidly 3.8–4.0 billion years ago in a hot, oxygen-free environment. This alien environment may or may not have been on Earth: terrestrial organisms have, after all, been found in extreme environments such as the Antarctic dry valleys, seething ocean vents and areas deep in the Earth.

Authors rushed to explain the new insights. Proliferating nearly as rapidly as the identifications of extrasolar planets themselves, a spate of popular books on planet detections appeared. Within 18 months of the initial discoveries, the searches had been described in the words of the scientists involved by Ken Croswell (*Planet Quest: The Epic Discovery of Alien Solar Systems*, Free Press). And the various techniques that ultimately yielded success were reviewed by Donald Goldsmith (*Worlds Unnumbered: The Search for Extrasolar Planets*, University Science Books) and Paul Halpern (*The Quest for Alien Planets: Exploring Worlds Outside the Solar System*, Plenum).

Following the distinguished tradition established by I. S. Shklovskii and Carl Sagan in their classic *Intelligent Life in the Universe* (Holden-Day), planetary researchers and astronomers, including Armand Delsemme (*Our Cosmic Origins: From the Big Bang to the Emergence of Life and Intelligence*), Bruce Jakosky (*Strangers in the Night: A Brief History of Life on Other Worlds*), S. Ross Taylor (*Destiny or Chance: Our Solar System and its Place in the Cosmos*, all Cambridge University Press) and Paul Davies (*The Fifth Miracle: The Search for the Origin and Meaning of Life*, Simon & Schuster), summarized the recent discoveries about the origin of life and then discussed their implications. All authors except the last give a resounding 'yes' to the question, 'Is extraterrestrial life abundant?'

In *Worlds Without End*, John S. Lewis, a noted planetary scientist and author of both semi-popular books and technical texts, covers much the same territory as the above books, at the level of *Nature's* News and Views section. He elegantly describes the constituents of our Solar System as they are known today. But, as he himself admits, "the planets of the Solar System ... offer a very limited sample of reality, and are an inadequate guide to the range of possible worlds. Worlds unknown vastly outnumber the



Circles of creation

How do planets form? This model of planetary formation, to be found in Santa Cruz, California, represents a Jupiter-sized planet that has carved out a groove in a disk of gas and dust around a young star. The planet's presence is sending ripples through the nebula, which might lead to the formation of more worlds. *The Planets*, by David McNab and James Younger

(Yale University Press, \$35), in which this picture may be found, chronicles our planetary travels, and describes how our understanding of the Solar System has developed from the first star-gazers in ancient times to Galileo and today. Space-race archives have been plumbed, and the book contains pictures from the Apollo, Voyager, Pioneer and Viking missions.

worlds we know." So he outlines current ideas about the formation of our Solar System, and reviews the now-overwhelming evidence for the existence of planets about other stars. Lewis's unique contribution is to combine basic physics and chemistry with examples from within the Solar System to predict the likely nature of distant planets and their satellite systems, especially with regard to their potential habitability.

Not surprisingly, coming from a cosmochemist whose influential research three decades ago elucidated the gross distribution of compounds and mineralogy across our Solar System, Lewis believes that "we can do a better job [of understanding] alien chemistry than alien biology". In this vein, he uses thermodynamic arguments to constrain the chemical make-up of other planetary worlds.

Lewis's language is playful at times, as in his tutorial about stellar astrophysics, or when he invents "Earthissmos" and "Earthlets" (large and small Earths), "Europoids", "Titanoids", etc. to explore the planetary possibilities. As this book shows, it is downright fun to try to wonder "about life as we don't know it". Generally speaking, these places sound a lot like home, and their inhabitants much like us. That is to say, many of Lewis's arguments come, at least in part, from their similarity to the Solar System. One must question whether such an extrapolation is justified. Surely the clearest message from Voyager's gradual unveiling of the outer Solar System was that "terrestrial chauvinism is unwarranted and misleading in the Universe at large". Already, the unexpected orbital character of the detected extrasolar planets — as compared with predictions — screams out: never believe theorists, whose forecasts

are almost always drably unimaginative.

Cognizant of the critical part played by stellar class and age, Lewis argues that life is most likely to be found around single stars of spectral classes G and K, extending perhaps into the F and M ranges. Even population I, but not II, brown dwarfs may successfully host life-supporting objects. He accepts James Kasting's idea of a habitable zone, a region around an astronomical unit or so where planets with appropriate spins can have clement climates. He maintains that planets of roughly Earth size to several Earths will have the temperatures and oceans that are a prerequisite for life. Lewis also suggests that life might, however, be found on a Europa-like body orbiting a giant planet or a brown dwarf; in the former case, a broader range of stellar classes is acceptable.

The book contains occasional jokes and asides on topics such as ice-skating well outside its scope. After opening with a choppy, brief survey of science fiction and historical ideas about exobiology (which has been bettered elsewhere), the book is technical in content, has no tables (some would have been preferred to paragraphs dense with numbers), and no equations or diagrams. In fact, it is remarkably devoid of space images or space art, normally commonplace in such books.

Lewis ends by addressing the ingredients needed for life. He concludes that water, the best of all possible solvents, is the medium in which the requisite chemistry occurs for life to originate and be sustained. He also maintains that, even in other worlds, carbon compounds — ideal carriers of complex information — will form the backbone of biological molecules, the actual structural material

out of which life's blueprints are made. Or is this just another example of terrestrial chauvinism? □

Joseph A. Burns is in the Departments of Astronomy and of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853, USA.

More on life beyond Earth

Planetary Dreams

by Robert Shapiro

Wiley, £22.50, \$27.95

Energetic Victorians

The Science of Energy: A Cultural History of Energy Physics in Victorian Britain

by Crosbie Smith

University of Chicago Press: 1999. 385 pp. \$25, £19.95

John Meurig Thomas

Writing about Hermann von Helmholtz in *Nature* (1876–77), James Clerk Maxwell asserted: "The scientific importance of the principle of the conservation of energy does not depend merely on its accuracy as a statement of fact, nor even on the remarkable conclusions which may be deduced from it, but on the fertility of the methods founded on this principle. It gives us a scheme by which we may arrange the facts of any physical science as instances of the transformation of energy from one form to another. It also indicates that in the study of any new phenomenon our first inquiry must be, 'How can this phenomenon be explained as a transformation of energy?'"

In 1847, Helmholtz had published his acclaimed *Über die Erhaltung der Kraft* ("On the Conservation of Force"); at that time, the law of conservation of energy had not been enunciated. In the intervening years there was an abundance of confusion and debate. Individuals whom scientists and historians now revere felt it necessary to indulge in a strange amalgam of innuendo, intemperate accusation and occasional bursts of vituperation, so strongly did they feel about the correctness of their own views and the misconceptions or deceptions of others. It surprised me to discover that religious preferences and (mildly) xenophobic sentiments were also brought into the cockpit of debate. Ultimately, however, "the baleful giant of Force" (to use the author's words) was dislodged by the "rightful monarch, Energy".

Much of the content of this meticulously researched, fascinating, fluently presented (but by no means flawless) book traces the genesis and growth of thermodynamics, especially the emergence of its first law. It also outlines the lives and arguments of the various protagonists who brought forth the whole science of energy. It is a cultural history, right enough, and I was surprised to be

exposed to so much biblical exegesis and social comment. The doctrine of "the universal dissipation of mechanical energy" we are told, could be read by Presbyterian students "as fully compatible with a traditional vision of a fallen world which was subject, like unregenerate souls, to depravity and death"!

Scottish natural philosophers, including the towering William Thomson (Lord Kelvin), P. G. Tait and Maxwell, and Scottish engineers such as William Macquorn Rankine and Henry Fleeming Jenkin, rightly loom large in the landscape painted by Crosbie Smith. But rather too much is made of the differences that existed in Victorian Britain between the Anglican academic establishment, with its Oxbridge axis, and the radical dissenting entrepreneurs (among them James Prescott Joule) of the industrializing North. After all, Kelvin, Tait and Maxwell entered Peterhouse (which the author persists in calling "St Peter's College") as undergraduates and all of them flourished as Fellows in Cambridge.

It is intriguing to note that, at about the time when Helmholtz was preparing his monograph "On the Conservation of Force", 26-year-old Kelvin — newly appointed to the chair of Natural Philosophy in Glasgow and preparing his undergraduate lectures there — had read a new edition of Thomas Young's *Course of Lectures on Natural Philosophy and the Mechanical Arts* (first published in 1807) and focused on Young's use of the word "energy". He subsequently championed and sharpened the definition of this word. Sixty years later, after the law of con-

servation of energy had long been accepted, the aged Kelvin, in a letter to his friend Joseph Larmor, disparaged the eminent German physical chemist Wilhelm Ostwald: "I do not know if (he) knows that energy is a capacity for doing work".

I was rather surprised to learn what Tait, in 1892, said of Henri Poincaré's *Thermodynamique*, which he denounced as a text "which exhibits a lavish, almost reckless use" of mathematical power, rendering it of little value to the student. And it was unexpected that in his 1875 lecture at the Royal Institution, Lord Rayleigh should have judged Josiah Willard Gibbs' paper "On the Equilibrium of Heterogeneous Substances" to be "too condensed and too difficult for most, I might say all, readers". Mercifully Maxwell did not think so, as the author reminds us, and this was a key factor in helping the early physical chemists to take full advantage of the phase rule and chemical potential.

The introductory chapter of this book is the least satisfactory. It makes far too much of the fact that innovators of science 'stage-manage' their credibility. Few working scientists, unlike historians of science, will be persuaded by the thesis that they go about their research animated by a quest for credibility. There are recurrent references in this text to "ascending and descending spirals of credibility", "credibility cycles" and the like. Most practising scientists pursue their research because they want to know answers to key questions. True, they are conscious of building up their reputations, but that is incidental to uncovering new knowledge. In

New in paperback

Korolev: How One Man Masterminded the Soviet Drive to Beat America to the Moon

by James Harford

Wiley, £13.99, \$17.95

First published in 1997. "Harford has produced a competent, earnest biography of Korolev, based mostly on secondary sources, that does not rise much above the derivative eulogy. Nor does it help very much our understanding the larger dilemmas that surround space activity at the end of the twentieth century". Alex Roland, *Nature* 392, 143–145 (1998).

The Fifth Miracle: The Search for the Origin of Life

by Paul Davies

Penguin Press, £8.99

"Davies's book is a small miracle in itself. In a little more than 200 pages he pursues the Mother of All Questions — 'What is life?' — in a way that should be deeply satisfying to physicists, chemists and biologists alike, and he does this in a clear and potent style that should

make the book equally stimulating to non-scientists ... Davies has little patience with those who take for granted that as soon as there are Earth-like conditions, life will sprout. This may be true, he says, but then we're making a gigantic assumption that should not be taken lightly. To explore this assumption, he takes the reader on an intellectual joyride up and down the entropy slope, along the way pointing his sharp flashlight at often murky concepts, such as order, organization, chance, randomness, specificity, language and semantics." Stefan Bengtson, *Nature* 395, 560 (1998).

Silent Spring

by Rachel Carson

Penguin, £7.99

The Edge of the Sea

by Rachel Carson

Penguin, £7.99

Fevered Lives: Tuberculosis in American Culture since 1870

by Catherine Ott

Harvard University Press, \$16.95

the introduction also, Thomas Kuhn's famous 1959 paper on "Energy Conservation as an Example of Simultaneous Discovery" is criticized largely because Kuhn adopts the stance of the omnipresent narrator rather than playing the role of the contextual historian — should not all good history be contextual? Smith's thesis I found unconvincing.

I also reject his claim that the Universe is fundamentally mechanical in nature. Has he forgotten that Michael Faraday's work on electrolysis, in the 1830s, uncovered the profound fact that electricity comes in units: in atoms. Matter and electricity are inextricably linked. As Richard Feynman once said of this particular discovery by Faraday, it was "one of the most dramatic moments in the history of science, one of those rare moments when two great fields come together and are unified". I was also disappointed that William Grove's work, which gave us the fuel cell, a vitally important means of generating power and energy from matter, is not mentioned. □

John Meurig Thomas is at the Master's Lodge, Peterhouse, University of Cambridge, Cambridge CB2 1QY, UK.

Thinking ahead (and why politicians don't)

Brain Policy: How the New Neuroscience Changes Our Lives and Our Politics

by Robert H. Blank

Georgetown University Press: 1999. 199 pp. \$60 (hbk), \$21.95 (pbk)

Joseph T. Coyle

Government policy is typically formulated by middle-aged bureaucrats and politicians. Experts may be consulted to present the current status of science, but this input is filtered through the policymakers' knowledge base — acquired during their education a quarter of a century or more before. Thus, decisions related to brain and behaviour (affecting policies on substance abuse, mental illness, neurodegenerative disorders and education) are often viewed through the prism of the 1970s, when deviant behaviour was explained by Freudian theory and brain disorders were accepted as incurable tragedies.

Brain Policy is Robert H. Blank's attempt to correct the viscosity of attitude change in response to scientific advances in understanding the brain. A social scientist, Blank became engaged in these issues through his membership between 1987 and 1991 of the US Office of Technology Assessment (OTA) Advisory Panel of Neuroscience Research. The OTA, which performed the important task of informing US government policymakers about scientific advances, was then eliminated as an "unnecessary expense". At least this minimized the confusion that new

Getting a grip on small things

This louse and its offspring hanging onto human hair can be found, along with a host of other oversized creatures, cells, crystals and micromachines, in *The Usborne Complete Book of the Microscope* by Kirsteen Rogers, edited by Paul Dowswell (Usborne, £9.99). The book won this year's Rhône-Poulenc Junior Science Book Prize and, besides its many revealing pictures, it contains a history of the microscope and instructions on how children can access the world of the microscope for themselves.



knowledge might have brought to old attitudes in formulating national policy.

Blank is justifiably energized by the remarkable and accelerating advances made in research on brain and behaviour. For example, the Society for Neuroscience grew over 25 years from 400 to 25,000 members in 1996: interest in neuroscience was exploding as the pace of discovery accelerated. Consistent with this growth, the US National Institutes of Health's annual commitments to neuroscience and behavioural research has increased from several hundred million to approximately \$4 billion over the same period. Not to be ignored is the recent finding that brain disorders account for most of the top ten health conditions, in terms of their economic burden on society.

Brain Policy has some real strengths and a few shortcomings. It is not contaminated by the special pleading of a neuroscientist. Nevertheless, converts can sometimes be more enthusiastic than the original members of a group, a problem Blank avoids by maintaining a critical distance from the *Zeitgeist* of neuroscience, pointing out the limits and dangers of "excessive exuberance" about therapeutic implications.

The opening chapters — explaining our current understanding of the development, organization and function of the human brain and the interaction among genes, neurons and environment — provide a first-rate introduction to neuroscience for the lay person. Blank condenses remarkably complex material into concepts that are readily grasped, but avoids gross oversimplification.

In later chapters he addresses some of the important policy controversies related to neuroscientific advances, including the current understanding of drug abuse and of behavioural genetics, the ethics of neural

grafting and the role of neurotoxins. He artfully side-steps simplistic genetic determinism of behaviour, emphasizing that serious psychiatric disorders are likely to involve complex interactions of several genes of risk with environment. He presents a nuanced discussion of the difficult ethical issues surrounding imposed treatment of psychiatric disorders. He justifiably criticizes the funding agencies for neglecting the study of neurotoxins' contribution to brain disease.

Occasionally, Blank falls into the trap that often occurs in arts and letters, wherein all assertions must be given equal weight. For example, he cites clinical reports that Prozac increased the risk of suicide and violence in depressed patients. While he indicates that the Food and Drug Administration rejected these findings, he does not make it clear that this decision was based on the results of controlled studies that disproved these unfounded contentions. This episode could have provided an instructive example of the distinction between scientific evidence and personal anecdotes; the latter often unfortunately shape policy and legal decisions. But these lapses are infrequent and do not detract from the overall thrust of the book.

Clearly the book is aimed at a general audience, but scientists not conversant with neuroscience would find it an informative, easy read. The first half of the book is redundant to the neuroscientist but the second half — covering policy implications, unintended consequences and ethical concerns — could prove illuminating. For ultimately, in the world of *realpolitik*, many elements such as ethics and culture play a role: policy is not determined by scientific facts alone. □

Joseph T. Coyle is in the Department of Psychiatry, Harvard Medical School, Boston, Massachusetts 02115, USA.

Directional guidance of neuronal migration in the olfactory system by the protein Slit

Wei Wu*, Kit Wong†, Jin-hui Chen*, Zhi-hong Jiang‡, Sophie Dupuis‡, Jane Y. Wu‡ & Yi Rao†

* Laboratory of Molecular Neurobiology, Shanghai Research Center for the Life Sciences and The National Institute of Neuroscience, The Chinese Academy of Sciences, 320 Yue-yang Road, Shanghai, China

† Department of Anatomy and Neurobiology, ‡ Departments of Pediatrics, and Molecular Biology and Pharmacology, Washington University School of Medicine, Box 8108, 660 S. Euclid Avenue, St Louis, Missouri 63110, USA

Although cell migration is crucial for neural development, molecular mechanisms guiding neuronal migration have remained unclear. Here we report that the secreted protein Slit repels neuronal precursors migrating from the anterior subventricular zone in the telencephalon to the olfactory bulb. Our results provide a direct demonstration of a molecular cue whose concentration gradient guides the direction of migrating neurons. They also support a common guidance mechanism for axon projection and neuronal migration and suggest that Slit may provide a molecular tool with potential therapeutic applications in controlling and directing cell migration.

Most neurons in the developing nervous system have to migrate from their birthplaces to reach their final positions. Studies of neuronal migration are important for revealing the mechanisms underlying the formation of a normal nervous system, for understanding the aetiology of human diseases caused by abnormal migration and for designing therapeutic approaches to neurological diseases.

Neuronal migration in the developing central nervous system (CNS) was initially inferred from histological observations by classical neuroembryologists including Vignal (1888), Ramon y Cajal (1891, 1911), Kolliker (1896), His (1904) and Hardesty (1904)¹ (reviewed in refs 2–5). The possibility that cells truly migrate, rather than that cells formed earlier are simply displaced by cells formed later⁶, was supported initially by cells moving in a direction opposite to that expected from cell displacement through histological examinations in the spinal cord of chick embryos⁷ and later by autoradiographic tracing in the cerebrum of rodent embryos⁸. The fact that only nuclei were traced in autoradiographic studies raised the possibility that nuclei, but not entire cells, moved in the highly structured nervous system. Electron microscopy and reconstruction provided strong evidence for the migration of neuronal cell bodies^{3–5}, and observations of primary neurons and glia cultured *in vitro* demonstrate directly that neurons do migrate^{9–10}. Using histological, autoradiographic, retroviral tracing, dye labelling and modern imaging techniques, it has now been established that the majority of neurons migrate throughout the developing nervous system^{11–12}.

Proper migration of neurons is essential for the formation and normal functioning of the nervous system. In humans, defects in neuronal migration can cause several diseases including epilepsy^{13,14}. Migration is also important for metastasis and invasion of neuroblastoma and glioma. Although it has been known for some time that cell migration is essential for postnatal behaviour changes in birds, we now know that neurogenesis and neuronal migration also continue in the brains of postnatal mammals, including humans. These findings indicate that, for successful application of cell-based therapies for neurodegenerative diseases, it is essential to direct the correct migration of neurons or cells expressing therapeutic products to the target region.

Our understanding of the molecular mechanisms guiding neuronal migration is still limited. Genetic studies in humans and mice

have identified many molecules, deficiencies of which cause defects in neuronal migration¹⁵. Because most of these are intracellular molecules, it is unlikely that they act as guidance cues for migrating neurons. One molecule identified from the genetic studies is the secreted protein Reelin^{16–18}. Loss-of-function mutations in the *reeler* gene cause defects in CNS lamination, and Reelin has been thought to control cell–cell interactions critical for cell positioning. It is not clear, however, whether Reelin acts as a stop signal or an adhesive molecule for migrating neurons^{15–18,21}. Because molecules can regulate neuronal migration indirectly²², it is not established whether Reelin acts directly or indirectly^{15,21}. Similarly, the precise roles of the other genetically identified molecules involved in neuronal migration remain to be determined^{15,21}.

The olfactory system is a useful model for studying neuronal migration. The olfactory bulb relays olfactory information from the olfactory epithelium to the primary olfactory cortex²³. The major types of interneuron in the olfactory bulb, including the granule cells and the periglomerular cells, are produced postnatally from the anterior subventricular zone (SVZa) of the telencephalon in rodents^{24–26}. Neuronal progenitors thus have to migrate in the rostral migratory stream (RMS) from the SVZa to the olfactory bulb^{19,20,27–32}. Co-culture of explants in collagen gel matrices indicated that the septum, at the midline of the telencephalon and caudal to the SVZa, secretes a repulsive factor(s) guiding the migration of SVZa neurons³¹. However, experiments involving a different culture method using matrigel found no repellent activity in the septum for migrating SVZa cells³². Because the matrigel allowed neurons to migrate on top of each other, a behaviour termed chain migration and thought to occur *in vivo*^{27,28,32}, the significance of the repulsive cue in the septum for guiding SVZa neurons was doubted by Wichterle *et al.*³². Furthermore, the molecular identity of the putative repellent(s) in the septum is not known.

We show here that the septum is repulsive to SVZa neurons in matrigel. We found the expression of two *slit* genes in the postnatal septum, encoding proteins that are repulsive cues capable of guiding the migration of SVZa cells. This guidance depends on a concentration gradient, rather than the absolute amount, of the Slit proteins. Moreover, Slit repels neurons migrating from the SVZa into its natural pathway, the RMS. These results indicate that Slit can act as a diffusible molecule guiding the direction of cell migration in the nervous system.

Repulsive activity in the septum

The existence of a repulsive activity in the septum for SVZa neurons was indicated by co-cultures of the SVZa and septum in collagen gel matrices³¹. Another assay for studying migration of SVZa neurons involves matrigel, a three-dimensional extracellular matrix gel of collagen IV, laminin, heparan sulfate proteoglycans and entactin-nidogen³³. The morphology and behaviour of SVZa cells cultured in the matrigel are thought to resemble those *in vivo*^{27,28,32}.

To test whether the septum is repulsive to SVZa neurons, we isolated postnatal SVZa and septal explants and co-cultured them in matrigel and collagen gel matrices. The distribution of neurons migrating out of the SVZa explants can reveal repulsive or attractive activities. When SVZa explants were placed in matrigel, the distribution of migrating cells was symmetric around the circumference of each explant (Fig. 1a). In the presence of the septum, more cells were found in the quadrant distal to the septum than those in the quadrant proximal to the septum (Fig. 1b) (36/36 explants). Thus, as in collagen gel matrices³¹ (Fig. 1d), the septum is repulsive to SVZa neurons. The ventral midline structure in the spinal cord, the floorplate, was also repulsive to the SVZa in matrigel (Fig. 1c) (12/12 explants). These results indicate that, although cells migrate on top of each other in matrigel^{27,28,32}, chains of migrating cells do respond to guidance cues in the ventral midline of the neural tube, including the septum and the floorplate.

Repulsion of migrating neurons by Slit

Genes encoding the secreted Slit proteins, which are chemorepellents for axons^{34–38}, are expressed in the embryonic septum^{34,35}. We have now found that *slit* genes are expressed in the postnatal murine septum. Of

the three *slit* genes, *slit-1* (*Slit-1*) and *slit-2* (*Slit-2*) are expressed in the postnatal septum and the neocortex (Fig. 1e–g), raising the possibility that Slit may be a guidance cue for SVZa neurons.

To investigate whether Slit proteins affected migration of neurons from the SVZa, we first co-cultured SVZa explants and Slit-expressing cells in matrigel. When SVZ explants from postnatal day 5 (P5) mice were co-cultured with human embryonic kidney (HEK) cells stably transfected with a control plasmid³⁴, chains of migrating neurons were symmetrically distributed (Fig. 2a) (32/32 explants). When SVZ explants were co-cultured with HEK cells stably expressing the *Xenopus* Slit protein (xSlit), an orthologue of mSlit-2, chain migration was highly asymmetric (Fig. 2c) (42/42); there were more chains in the quadrant distal to the Slit cells than in the quadrant proximal to the Slit cells. The same effect occurred when SVZa explants were co-cultured with mSlit-1-expressing cells (Fig. 2e) (16/16). These results indicate that Slit is repulsive for SVZ cells migrating in chains.

We have also assayed SVZa migration using collagen gel matrices³¹. Although the distribution of migrating SVZa cells was symmetric when explants were co-cultured with control cells (Fig. 2b) (104/122 explants), with xSlit cells there were more SVZa cells in the distal quadrant (Fig. 2d) (235/254 explants). Cells transiently expressing mSlit-1 have similar effects (Fig. 2f) (31/31). The effect of xSlit on SVZa cells was quantified, showing that cells in the distal quadrant were not only more numerous, but also farther away from the explants (Fig. 3d). Because single cells, rather than chains of cells, migrate in collagen gel matrices, the repulsive activity of Slit in the collagen gel assay indicates that Slit can act on single cells.

To find out whether the migrating cells were neurons, we carried out immunocytochemistry with the TuJ1 antibody, which recognizes the neuron-specific β -tubulin (Fig. 3). The migrating cells stained positive for TuJ1, confirming that they are neurons

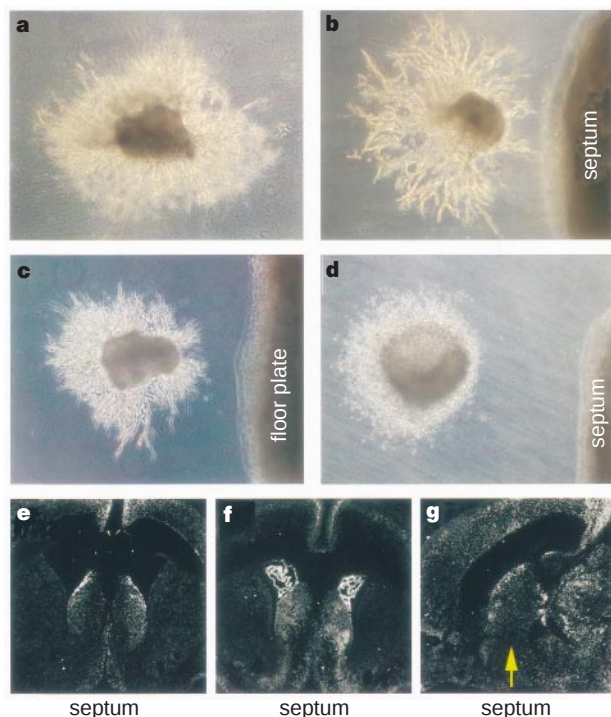


Figure 1 Repulsion of SVZa neurons by the septum and the floorplate and expression of *slit* genes in the postnatal septum. **a**, Chains of cells migrate out of SVZa explants symmetrically when cultured alone in matrigel. In the presence of the septum (**b**) or floorplate (**c**), there are more migrating cells in the quadrant of the SVZa explant distal to the septum than in the proximal quadrant. **d**, As reported³¹, the septum is repulsive to SVZa cells in collagen gel. **e**, A coronal section showing expression of *Slit-1* in the septum and the neocortex of postnatal day 3 (P3) rats. **f**, A coronal section showing expression of *Slit-2* in the septum, the choroid plexus and the neocortex of P3 rats. **g**, A sagittal section showing expression of *Slit-1* in the septum and the neocortex of P3 rats. Rostral is to the left and dorsal is up.

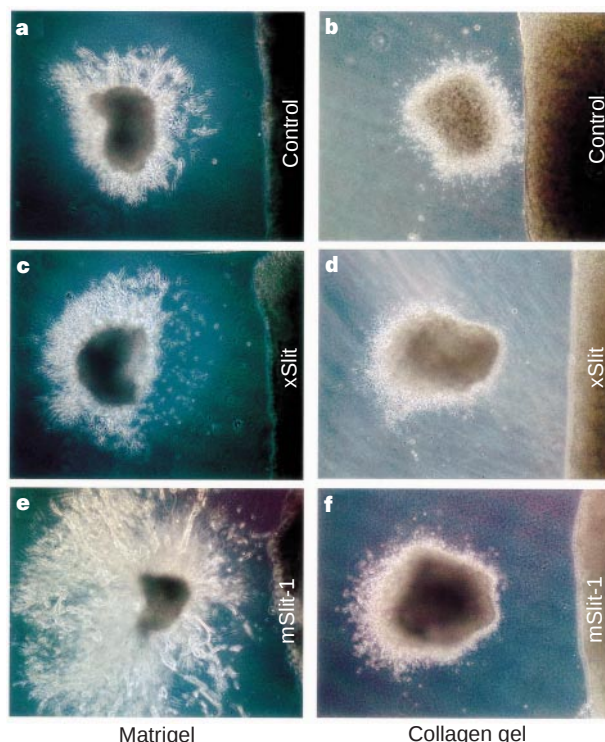


Figure 2 Effect of Slit on neurons migrating from SVZa explants. **a**, **c**, **e**, Distribution of chains of cells from the SVZa explants, co-cultured with an aggregate of control HEK cells (**a**), with cells expressing xSlit (**c**) or mSlit-1 (**e**) in matrigel for one (**a**, **c**) or two days (**e**). **d**–**f**, Distribution of cells migrating out of the SVZa explants after being co-cultured with control HEK cells (**b**), with cells expressing xSlit (**d**) or mSlit-1 (**f**) in collagen gel.

(Fig. 3a–c). The uneven distribution of TuJ1 staining within the SVZa explants indicates that neurons migrated within each explant away from the Slit cells (Fig. 3c).

These findings could be explained in two ways: either Slit repels neurons, or Slit inhibits migration. To distinguish between them, we placed two SVZa explants at different distances from a single aggregate of Slit cells. The distance from the Slit cells to the proximal side of the more distant explant was equal to, or slightly farther than, the distance from the Slit cells to the distal side of the closer explant (Fig. 4b). In both explants there were more cells in the distal quadrants. These results are best explained by a repulsive activity of Slit, rather than inhibition of migration by Slit. Next we placed an SVZa explant on top, rather than to the side, of Slit-expressing cells. Neurons still migrated out of the SVZa (Fig. 4d) (16 explants), arguing against inhibition of neuronal migration by Slit. These results indicate that Slit is repulsive to, rather than inhibitory of, migrating neurons.

To test whether SVZa cells respond to the absolute concentration of Slit or to a concentration gradient of the protein, we placed SVZa explants between two aggregates of Slit cells. When an SVZa explant was placed at unequal distances from two Slit aggregates, neurons migrated away from the closer Slit aggregate (Fig. 4e). When SVZa explants were placed at an equal distance from two Slit aggregates, the cells migrated out symmetrically (Fig. 4f). The simplest explanation for these observations is that SVZa cells respond to a concentration gradient of the Slit protein.

To determine the effective distance of Slit, multiple SVZa explants were placed at different distances from control or Slit cells. When the explants were within 1 mm of the Slit cells, neurons migrated asymmetrically (Fig. 4h).

Slit and neuronal migration in the RMS

Although Slit can repel neurons migrating out of SVZa explants in collagen gels and in matrigel, the question of whether Slit can guide neurons migrating in their natural pathway was not addressed. We developed an assay to investigate the effect of Slit on neurons migrating from the SVZa into their natural pathway, the RMS, towards the olfactory bulb. Sagittal sections of postnatal brains

containing the SVZa, the RMS and the olfactory bulb were isolated and cultured. Crystals of the lipophilic dye 1,1'-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate (DiI) were placed into the SVZa to label neuronal precursors. After 24 h in culture, migrating neurons were found in the RMS (Fig. 5b). To test the effect of Slit, control HEK cells or HEK cells stably expressing Slit were pre-labelled with another lipophilic dye 3,3'-dioctadecyloxa-carbocyanine (DiO), and placed on top of the RMS. With control HEK cells, SVZa neurons migrated into the RMS (Fig. 5a–c) (14/14 explants). In contrast, when Slit cells were placed on the RMS, few SVZa neurons migrated into the RMS (Fig. 5d–f) (16/16). These results provide strong evidence indicating that Slit can regulate neuronal migration in natural pathways.

Role of endogenous Slit in the septum

The septum contains a repulsive activity for SVZa neurons³¹. We have also found that *slit* genes are expressed in the septum and that Slit proteins are repulsive to SVZa neurons. Together, these results indicate that endogenous Slit may be involved in repelling SVZa neurons. To test directly whether endogenous Slit contributes to the repulsive activity in the septum, we constructed a plasmid that

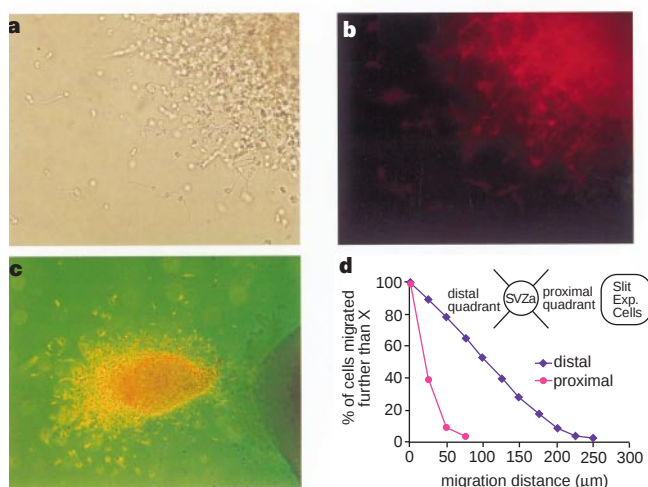


Figure 3 Neuronal nature and distribution of SVZa neurons. **a**, Cells migrating out of the SVZa explant. **b**, The same cells as in **a**, stained with the TuJ1 antibody. Under different focal planes, essentially all cells are TuJ1 positive. **c**, Superimposition of bright light and fluorescent views of another explant co-cultured with Slit-expressing cells. Note the asymmetric distribution of neurons within the SVZa explant. **d**, Quantification of the effect of Slit on migrating SVZa neurons after TuJ1 staining. X axis, distance of neurons from the edge of SVZa explants; Y axis, percentage of neurons that have migrated further than the distance indicated on the X axis. Similar results have been obtained from five experiments.

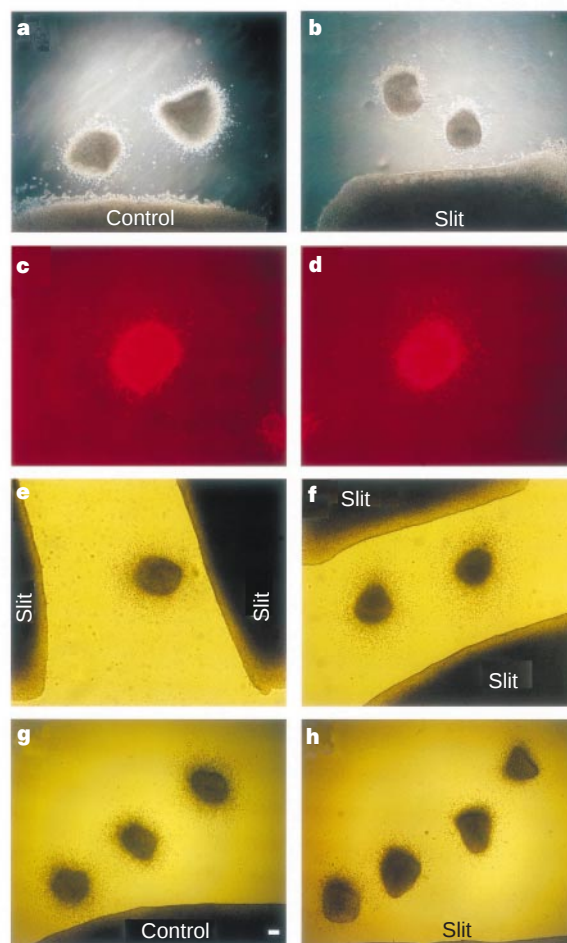


Figure 4 Spatial relationship between SVZa explants and Slit-expressing cells. Results of co-cultures in collagen gels are shown. **a, b**, SVZa explants placed next to HEK cells. **c, d**, SVZa explants placed on top of HEK cells and migrating neurons revealed by staining with the TuJ1 antibody. **e**, When a single SVZa explant was placed between two aggregates of Slit cells, the asymmetric distribution of cells shows that neurons migrated according to guidance from the closer aggregate. **f**, When two SVZa explants were placed at an equal distance from two aggregates of Slit cells, the distribution of SVZa neurons is symmetric. **g, h**, Placement of multiple SVZa explants allows determination of the effective distance of Slit, with the most distal explant at ~1 mm.

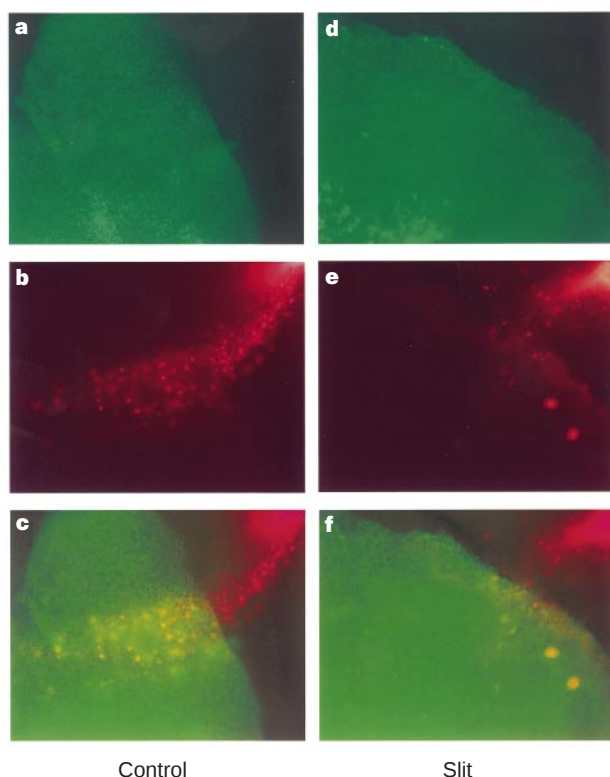


Figure 5 Migration of neurons in the RMS. Dil crystals were placed into SVZa in sagittal sections of postnatal rat brains to label cells (red in **b, c, e, f**) migrating into the RMS. Control or Slit HEK cells were labelled with DiO (green in **a, c, d, f**). **a–c**, different views of the same slice in which an aggregate of control HEK cells were placed on top of the RMS. **d–f**, Different views of the same slice in which an aggregate of Slit cells were placed on top of the RMS. In all panels, the upper right corner is towards the SVZa, the origin of the RMS, whereas the lower left corner is towards the olfactory bulb, the end of the RMS.

expressed RoboN, the extracellular domain of Roundabout (Robo), a receptor for Slit^{34,36–38}. After complementary DNA transfection into HEK 293T cells, RoboN protein was secreted into the medium and could bind to the Slit protein (Fig. 6c). Because RoboN lacks the transmembrane and intracellular domains, it could be a competitive inhibitor of Slit–Robo signalling.

To investigate whether RoboN could inhibit the activity in the septum, we placed explants of the septum on top of either control HEK cells or cells expressing RoboN, and then co-cultured them with explants of the SVZa, and examined the distribution of migrating SVZa neurons. When cultured on top of the control cells, the septum was effective in repelling SVZa neurons (Fig. 6a). RoboN cells significantly reduced the number of explants with asymmetric distribution of migrating neurons (Fig. 6b, d). Although these results did not distinguish which endogenous Slit was involved, they indicate that endogenous Slit contributes to the repulsion of SVZa neurons by the septum.

Discussion

We have shown that Slit is a chemorepellent for neurons migrating from the SVZa and that Slit acts as a diffusible molecule, the concentration gradient of which guides the direction of migrating neurons. Slit is, to our knowledge, the first molecule directly shown to be sufficient for the directional sensing of migrating neurons. With previous findings that Slit is an axon repellent^{34–38}, these results also support the idea that there are at least some molecular guidance mechanisms in common between axon projection and neuronal migration.

Slit is a secreted protein^{34–38}. Our data indicate that Slit can guide

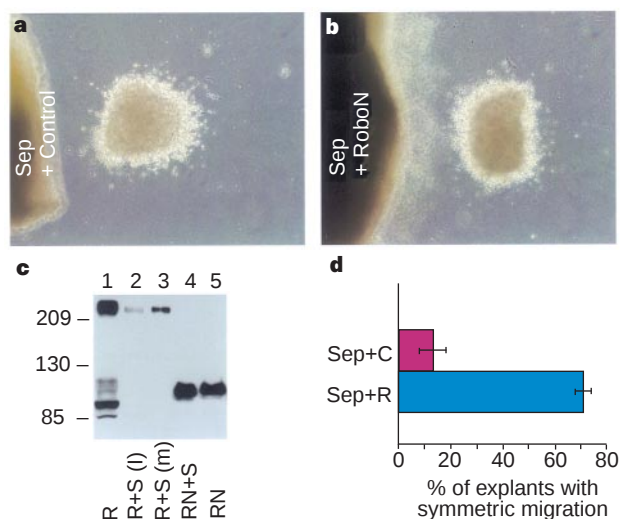


Figure 6 Inhibition of the repulsive activity in the septum by RoboN. **a, b**, Distribution of SVZa neurons after co-culturing with septal explants on top of either control HEK cells (**a**) or HEK cells expressing RoboN (**b**). **c**, Binding of RoboN to Slit. Numbers on the left indicate relative molecular mass. HA-tagged full-length Robo (lane 1) and RoboN (lane 5) were detected by an anti-HA antibody; Robo was co-incubated either with Myc-tagged xSlit lysate (lane 2) or with Myc-tagged xSlit in the medium (lane 3) and Robo was detected by the anti-HA antibody after immunoprecipitation with the anti-Myc antibody. Lane 4: RoboN was co-incubated with Myc-tagged xSlit in medium and detected by the anti-HA antibody after immunoprecipitation with the anti-Myc antibody. **d**, Percentage of SVZa explants with symmetric migration. Sep + C, results with the septal explants cultured on top of control cells ($n = 36$), Sep + R, results with the septal explants cultured on top of RoboN cells ($n = 73$). The difference between these two groups was statistically significant ($P < 0.05$ in Student's t -test).

neuronal migration without contact between the cells responding to Slit and those producing Slit. It is unlikely that Slit acts indirectly by regulating the differentiation of HEK cells or cells in the SVZa explants. The repulsion of individual neurons by Slit in collagen gels further supports the idea that Slit acts directly on migrating neurons. Our studies have shown that Slit is a repulsive cue, rather than an inhibitory factor, for migrating neurons. The behaviour of SVZa explants at different distances from two Slit sources indicates that the concentration gradient of Slit is likely to be responsible for directional sensing. The ephrins can prevent neural crest cells from migrating on caudal somites^{39–41}. Although the fact that ephrins are membrane-anchored proteins makes it unlikely that they act as diffusible guidance cues, it remains to be seen whether Slit and ephrins share any molecular mechanisms.

In the olfactory system, interneuron precursors migrate several millimetres in the RMS from the SVZa to the olfactory bulb. The activities of the septum and Slit indicate that Slit should be able to guide at least the initial migration of cells from the SVZa. It is not clear whether a gradient of Slit is responsible for guiding neuronal migration in the entire RMS. In our experiments with SVZa explants, the Slit cells could not act at distances greater than 1 mm. However, it cannot be ruled out that an endogenous Slit gradient could be laid down by the septum over a longer time over the entire RMS. It is also possible that other cues, acting in or around the RMS, may affect the migration of SVZa cells. Because significant influence over the direction of SVZa cell migration was only exerted by the septum, but not by the target tissue (the olfactory bulb or by tissues surrounding the RMS³¹, it seems reasonable to suppose that, if other molecules help to guide SVZa migration, they may act not just by themselves, but in coordination with Slit. Molecules such as the neural cell adhesion molecule (NCAM), which is expressed in the RMS and is necessary for SVZa migration^{30,42,43}, may act locally in the migration pathway to

allow adhesion of SVZ neurons. It will be interesting to test whether there is any functional relationship between Slit and NCAM. The embryonic septum can repel embryonic olfactory bulb axons and neonatal SVZa neurons, whereas the neonatal septum can repel only migrating neonatal SVZa neurons but not embryonic olfactory bulb axons³¹. The expression and activity of *slit* in both embryonic and neonatal septum indicate that either different forms of Slit proteins may be responsible for axon guidance and neuronal migration, or there may be other molecules modulating the activity of the same Slit proteins. Although only the caudal, not the rostral, septum has been shown to repel SVZa neurons³¹, our *in situ* hybridization has not revealed obvious differences in the expression of *slit* genes in the rostral and caudal septum, again raising the possibility of other modulating activities.

Functional studies of Slit make it interesting to consider the relationship between neuronal migration and axon guidance. Both processes involve cell motility, but the entire cell body moves in one case, whereas only axons move in the other case. Defects in both axon pathways and neuronal position can be caused by mutations in the same genes. Loss-of-function mutations in *netrin*, its *Caenorhabditis elegans* homologue *unc-6* or their receptors result in abnormal axon projection and cell position in the mouse^{44–46} and *C. elegans*⁴⁷. Genetic and phenotypic analyses in *C. elegans* indicate that *unc-6* is required for the migration of mesodermal cells⁴⁷. Experiments with commissural and other explants have demonstrated that Netrin-1 can directly guide axon projections⁴⁴, and recent evidence from explant studies indicates that defects in pontine neuronal migration in mouse mutants similarly reflect a direct attractive action of Netrin-1 on these neurons (K. Lee, H. Simon, M. Tessier-Lavigne and D. O'Leary, personal communication). It will be interesting to see if the same is true of other cell migration defects reported in mice mutant for Netrin or Netrin receptors, or whether some of them are due to indirect effects on the projection of axons or glial fibres along which neurons may migrate (see ref. 48 for review of neurophilic and gliophilic migration).

Genetic and explant experiments indicate that Slit acts directly as a chemorepellent for axons^{34,35,37,38}. Mutations in *robo* and *slit* in *C. elegans* and *Drosophila* result in defective cell positioning, suggesting that Slit–Robo signalling is required for cell migration^{37,49}. Genetic and phenotypic analysis in *Drosophila* suggests that *slit* is required for the migration of muscle precursor cells³⁷; certain muscle precursor cells, which normally migrate away from the embryonic midline, do not migrate in *slit* mutants (T. Kidd and C. S. Goodman, personal communication). Our present results have shown directly that Slit is sufficient to act as a repellent for migrating neurons. Slit has therefore been directly demonstrated to guide both neuronal migration and axon projection^{34–38}. It seems that whether Slit guides the movement of cell bodies or axons depends on the type of responding cell. It is unclear, however, whether intracellular machinery or extracellular mechanisms determine the type of response. Is the same form of Slit protein involved? Are there more intracellular components to enable a neuron to move its cell body? Or does the absence of critical components cause a neuron to move its cell body rather than its axon? The availability of a guidance cue acting *in vitro* on both migrating neurons and projecting axons offers the prospect of addressing these questions experimentally.

Cell migration is crucial in other developmental and non-developmental processes, including migration of cortical neurons radially along glial fibres, migration of neural-crest precursor cells, gastrulation, heart formation, muscle development, angiogenesis, leukocyte chemotaxis and tumour metastasis. We have obtained evidence for Slit function in other populations of neurons⁵⁰, indicating that guidance cues may be crucial in many processes involving neuronal migration. It will therefore be interesting to test whether the Slit family or other guidance cues can guide migrating cells in multiple systems in normal and abnormal situations. For migrating cells that do not respond to Slit, it will be

interesting to investigate whether the introduction of Robo or other potential receptor components can confer responsiveness to Slit. This may broaden potential therapeutic applications of Slit proteins in controlling unwanted cell migration and in delivering cells to the intended target region. □

Methods

In situ hybridization. Brains were removed from P3 rats, and 16- μ m coronal sections were cut with a cryostat, collected on superfrost plus slides and air dried. ³⁵S-labelled *slit* probes were made and 2×10^6 counts per minute (c.p.m.) of riboprobes was applied to each slide with brain slices. The slides were overlaid with coverslips and incubated in humidified chambers at 55 °C overnight before being washed with a washing buffer (50% formamide, $2 \times$ SSC and 0.1% β -mercaptoethanol) at 65 °C for 30 min, rinsed for 10 min in RNase buffer (0.5 M NaCl, 10 mM Tris-HCl, pH 7.5, 5 mM EDTA) and treated for 30 min at 37 °C with $20 \mu\text{g ml}^{-1}$ RNase A. The slides were then placed at 65 °C in the washing buffer for 5 min at 37 °C in $2 \times$ SSC for 15 min and at 37 °C in $0.1 \times$ SSC for 15 min before being dehydrated. The slides were coated with Kodak NTB2 emulsion and stored at 4 °C for 3 weeks before being developed in Kodak D19 and counterstained with haematoxylin.

Co-culture of explants. Brains from newborn Sprague-Dawley rats (postnatal days 3–7) were embedded in 7% low melting-point agarose prepared in phosphate-buffered saline. Coronal and sagittal sections of 400 μ m were cut with a vibratome. Tissue within the borders of the SVZ in coronal sections was dissected out to make SVZa explants of 200–400 μ m in diameter. We isolated septal explants from sagittal sections and placed them into the collagen gel or matrigel as described^{31,32,34}. They were cultured for 16–24 h under 5% CO₂ in F-12 medium supplemented with 10% fetal bovine serum, and penicillin and streptomycin.

Co-culture of SVZa explants with cell aggregates. We have two stable HEK cell lines³⁴. One was transfected with a plasmid expressing xSlit, and the other with the vector plasmid. Both lines went through similar selection and subcloning. The Slit line was confirmed by western blotting to express the full-length xSlit protein tagged with the Myc epitope. We also transiently transfected HEK 293 cells with either a vector or a plasmid expressing full-length Myc-tagged mSlit-1. The expression of mSlit-1 was also confirmed by western analysis. Because xSlit and mSlit-1 behaved similarly in all tests, we usually used the xSlit stable line.

We made aggregates of the control or Slit-expressing cells by the hanging-drop method. They were placed into collagen gels or matrigel with SVZa explants, and cultured as described above.

For immunohistochemistry, samples were fixed with 4% paraformaldehyde and pre-incubated with 10% goat serum in 0.1 M phosphate buffer (pH 7.4) for 1 h. Incubation with the primary antibody (TuJ1, 1:200 dilution) was carried out overnight at 4 °C. They were washed and then incubated with a goat anti-mouse secondary antibody conjugated to Cy3 for 1 h at room temperature, and washed before being mounted with glycerol.

For quantification, we obtained immunofluorescence images of samples stained with TuJ1, using a Zeiss microscope. We quantified cell distribution by analysing the distance between the cell body and the nearest edge of the explant, using the IPLab program.

To test whether the extracellular domain of Robo (RoboN) could inhibit the septal activity, aggregates were made from HEK cells transiently transfected with a plasmid expressing RoboN. These cells were embedded in a thin layer of collagen gel. Explants of the caudal septum were placed on top of the RoboN cells and co-cultured with SVZa explants in collagen gel matrices.

Slice assay with the RMS. Sagittal sections of postnatal rats were cut with a vibratome and those containing the SVZa, the RMS and the olfactory bulb were used. Crystals of DiI were inserted into the SVZa. The slices were cultured in a similar way to the SVZa explants.

Control or Slit-expressing HEK cells were labelled with DiO. Aggregates of these cells were placed on top of the RMS. We visualized DiI and DiO signals with different filters under the microscope. Images were taken with a Spot camera and stored on the computer.

Binding of RoboN to Slit. Full-length Robo has 1,660 residues; RoboN was made by tagging a hemagglutinin (HA) epitope to amino-acid residue 718. Lysates of cells expressing HA-tagged full-length Robo or RoboN were

incubated with lysates or media of cells expressing xSlit. Immunoprecipitation was done as described³⁴.

Received 8 March; accepted 8 June 1999.

1. Hardesty, I. On the development and nature of the neuroglia. *Am. J. Anat.* **3**, 229–268 (1904).
2. Cajal, S. R. *Histology of the Nervous System* (trans. Swanson, N. & Swanson, L. W.) (Oxford Univ. Press, New York, 1995).
3. Rakic, P. Neuron–glia relationship during granule cell migration in developing cerebellar cortex. *J. Comp. Neurol.* **141**, 283–312 (1971).
4. Rakic, P. Guidance of neurons migrating to the fetal monkey neocortex. *Brain Res.* **33**, 471–476 (1971).
5. Rakic, P. Mode of cell migration to the superficial layers of fetal monkey neocortex. *J. Comp. Neurol.* **145**, 61–84 (1972).
6. Tilney, F. Behavior in its relation to the development of the brain. Part II. Correlation between the development of the brain and behavior in the albino rat from embryonic states to maturity. *Bull. Neurol. Inst. NY* **3**, 252–358 (1933).
7. Levi-Montalcini, R. The origin and development of the visceral system in the spinal cord of the chick embryo. *J. Morphol.* **86**, 253–278 (1950).
8. Angevine, J. B. Jr & Sidman, R. L. Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. *Nature* **192**, 766–768 (1961).
9. Hatten, M. E. & Liem, R. H. K. Astroglia provide a template for the positioning of developing cerebellar neurons *in vitro*. *J. Cell Biol.* **90**, 622–630 (1981).
10. Hatten, M. E., Liem, R. H. K. & Mason, C. A. Two forms of glial cells interact differently with neurons *in vitro*. *J. Cell Biol.* **98**, 193–204 (1984).
11. Gray, G. E., Leber, S. M. & Sanes, J. R. Migratory patterns of clonally related cells in the developing central nervous system. *Experientia* **46**, 929–940 (1990).
12. Walsh, C. & Cepko, C. Cell lineage and cell migration in the developing cerebral cortex. *Experientia* **46**, 940–947 (1990).
13. Norman, M. G., McGillivray, B. C., Kalousek, D. K., Hill, A. & Poskitt, K. J. in *Congenital Malformations of the Brain: Pathologic, Embryological, Clinical, Radiological and Genetic Aspects* (ed. Norman, M. G.) 223–277 (Oxford Univ. Press, New York, 1995).
14. Gordon, N. Epilepsy and disorders of neuronal migration. II: Epilepsy as a symptom of neuronal migration defects. *Dev. Med. Child. Neurol.* **38**, 1131–1134 (1996).
15. Pearlman, A. L., Faust, P. L., Hatten, M. E. & Brunstrom, J. E. New directions for neuronal migration. *Curr. Opin. Neurobiol.* **8**, 45–54 (1998).
16. Falconer, D. S. Two new mutants 'trembler' and 'reeler', with neurological actions in the house mouse. *J. Genet.* **50**, 192–201 (1951).
17. D'Arcangelo, G. *et al.* A protein related to extracellular matrix proteins deleted in the mouse mutant *reeler*. *Nature* **374**, 719–723 (1995).
18. Hirotsune, S. *et al.* The *reeler* gene encodes a protein with an EGF-like motif expressed by pioneer neurons. *Nature Genet.* **10**, 77–83 (1995).
19. Luskin, M. B. Restricted proliferation and migration of postnatally generated neurons derived from the forebrain subventricular zone. *Neuron* **11**, 173–189 (1993).
20. Lois, C. & Alvarez-Buylla, A. Proliferating subventricular zone cells in the adult mammalian forebrain can differentiate into neurons and glia. *Proc. Natl Acad. Sci. USA* **90**, 2074–2077 (1993).
21. Frotscher, M. Cajal–Retzius cells, Reelin, and the formation of layers. *Curr. Opin. Neurobiol.* **8**, 570–575 (1998).
22. Rio, C., Rieff, H. I., Qi, P. & Corfas, G. Neuregulin and erbB receptors play a critical role in neuronal migration. *Neuron* **19**, 39–50 (1997).
23. Farbman, A. I. in *Smell and Taste in Health and Disease* (eds Getchell, T. V. *et al.*) 19–32 (Raven, New York, 1991).
24. Altman, J. & Das, G. D. Autoradiographic and histological studies of postnatal neurogenesis I: a longitudinal investigation of the kinetics, migration and transformation of cells incorporating tritiated thymidine in neonate rats, with special reference to postnatal neurogenesis in some brain regions. *J. Comp. Neurol.* **127**, 337–390 (1966).
25. Hinds, J. W. Autoradiographic study of histogenesis in the mouse olfactory bulb. II cell proliferation and migration. *J. Comp. Neurol.* **134**, 305–322 (1968).
26. Altman, J. Autoradiographic and histological studies of postnatal neurogenesis IV: cell proliferation and migration in the anterior forebrain, with special reference to persisting neurogenesis in the olfactory bulb. *J. Comp. Neurol.* **137**, 433–458 (1969).
27. Lois, C. & Alvarez-Buylla, A. Long-distance neuronal migration in the adult mammalian brain. *Science* **264**, 1145–1148 (1994).
28. Lois, C., Garcia-Verdugo, J.-M. & Alvarez-Buylla, A. Chain migration of neuronal precursors. *Science* **271**, 978–981 (1996).
29. Zigova, T. *et al.* A comparison of the patterns of migration and the destinations of homotopically transplanted neonatal subventricular zone cells and heterotopically transplanted telencephalic ventricular zone cells. *Dev. Biol.* **173**, 459–474 (1996).
30. Hu, H., Tomaszewicz, H., Magnuson, T. & Rutishauser, U. The role of polysialic acid in migration of olfactory bulb interneurons precursors in the subventricular zone. *Neuron* **16**, 735–743 (1996).
31. Hu, H. & Rutishauser, U. A septum-derived chemorepulsive factor for migrating olfactory interneuron precursors. *Neuron* **16**, 933–940 (1996).
32. Wichterle, H., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Direct evidence for homotypic, glia-independent neuronal migration. *Neuron* **18**, 779–791 (1997).
33. Kleinman, H. K., McGravey, M. L., Liotta, L. A., Robey, P. G., Tryggvason, K. & Martin, G. R. Isolation and characterization of type IV procollagen, laminin, and heparan sulfate proteoglycan from the EHS sarcoma. *Biochemistry* **21**, 6188–6193 (1982).
34. Li, H. S. *et al.* Vertebrate Slit, a secreted ligand for the transmembrane protein roundabout, is a repellent for olfactory bulb axons. *Cell* **96**, 807–818 (1999).
35. Nguyen Ba-Charvet, K. T. *et al.* Slit2-mediated chemorepulsion and collapse of developing forebrain axons. *Neuron* **22**, 463–473 (1999).
36. Wang, K. H. *et al.* Purification of an axon elongation- and branch-promoting activity from brain identifies a mammalian Slit protein as a positive regulator of sensory axon growth. *Cell* **96**, 771–784 (1999).
37. Kidd, T., Bland, K. S. & Goodman, C. S. Slit is the midline repellent for the Robo receptor in *Drosophila*. *Cell* **96**, 785–794 (1999).
38. Brose, K. *et al.* Evolutionary conservation of the repulsive axon guidance function of Slit proteins and of their interactions with Robo receptors. *Cell* **96**, 795–806 (1999).
39. Krull, C. E. *et al.* Interactions of Eph-related receptors and ligands confer rostrocaudal pattern to trunk neural crest migration. *Curr. Biol.* **7**, 571–580 (1997).
40. Smith, A., Robinson, V., Patel, K. & Wilkinson, D. G. The EphA4 and EphB1 receptor tyrosine kinases and ephrin-B2 ligand regulate targeted migration of branchial neural crest cells. *Curr. Biol.* **7**, 561–570 (1997).
41. Wang, H. U. & Anderson, D. J. Eph family transmembrane ligands can mediate repulsive guidance of trunk neural crest migration and motor axon outgrowth. *Neuron* **18**, 383–396 (1997).
42. Ono, K., Tomaszewicz, H., Magnuson, T. & Rutishauser, U. N-CAM mutation inhibits tangential neuronal migration and is phenocopied by enzymatic removal of polysialic acid. *Neuron* **13**, 595–609 (1994).
43. Rousselot, P., Lois, C. & Alvarez-Buylla, A. Embryonic (PSA) N-CAM reveals chains of migrating neuroblasts between the lateral ventricle and the olfactory bulb of adult mice. *J. Comp. Neurol.* **351**, 51–61 (1995).
44. Serafini, T. *et al.* Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**, 1001–1014 (1996).
45. Fazeli, A. *et al.* Phenotype of mice lacking functional *Deleted in colorectal cancer (Dcc)* gene. *Nature* **386**, 796–804 (1997).
46. Ackerman, S. L. *et al.* The mouse rostral cerebellar malformation gene encodes an UNC-5-like protein. *Nature* **386**, 838–842 (1997).
47. Hedgecock, E. M., Culotti, J. G. & Hall, D. H. The unc-5, unc-6, and unc-40 genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*. *Neuron* **4**, 61–85 (1990).
48. Rakic, P. Principles of neural migration. *Experientia* **46**, 882–891 (1990).
49. Zallen, J. A., Yi, B. A. & Bargmann, C. I. The conserved immunoglobulin superfamily member SAX-3/Robo directs multiple aspects of axon guidance in *C. elegans*. *Cell* **92**, 217–227 (1998).
50. Zhu, Y., Li, H. S., Zhou, L., Wu, J. Y. & Rao, Y. Cellular and molecular guidance of GABAergic neuronal migration from an extracortical origin to the neocortex. *Neuron* (in the press).

Acknowledgements. We thank W. Yuan and D. Ornitz for mouse *slit* cDNAs; J. Xu, Q. Wang and L. Zhou for help with *in situ* hybridization; W. Gan and J. Lichtman for help with confocal imaging and analysis; J. Brunstrom and A. Pearlman for discussions; C. S. Goodman and M. Tessier-Lavigne for comments; NIH, NSFC and SCST for support; and the John Merck Fund, NSFC and the Leukemia Society of America for scholar awards (to Y.R. and J.Y.W.).

Correspondence and requests for materials should be addressed to J.Y.W. or Y.R. (e-mail: jwu@pharmsun.wustl.edu; raoyi@thalamus.wustl.edu).

Reducing vortex density in superconductors using the 'ratchet effect'

C.-S. Lee*, B. Jankó†, I. Derényi‡ & A.-L. Barabási*

* Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA

† Materials Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA

‡ Department of Surgery, MC 6035, University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637, USA

A serious obstacle impeding the application of low- and high-temperature superconductor devices is the presence of trapped magnetic flux^{1,2}: flux lines or vortices can be induced by fields as small as the Earth's magnetic field. Once present, vortices dissipate energy and generate internal noise, limiting the operation of numerous superconducting devices^{2,3}. Methods used to overcome this difficulty include the pinning of vortices by the incorporation of impurities and defects⁴, the construction of flux 'dams'⁵, slots and holes⁶, and magnetic shields^{2,3} which block the penetration of new flux lines in the bulk of the superconductor or reduce the magnetic field in the immediate vicinity of the superconducting device. The most desirable method would be to remove the vortices from the bulk of the superconductor, but there was hitherto no known phenomenon that could form the basis for such a process. Here we show that the application of an alternating current to a superconductor patterned with an asymmetric pinning potential can induce vortex motion whose direction is determined only by the asymmetry of the pattern. The mechanism responsible for this phenomenon is the so-called 'ratchet effect'⁷⁻¹⁰, and its working principle applies to both low- and high-temperature superconductors. We demonstrate theoretically that, with an appropriate choice of pinning potential, the ratchet effect can be used to remove vortices from low-temperature superconductors in the parameter range required for various applications.

We consider a type II superconductor film of the geometry shown in Fig. 1a, placed in an external magnetic field H . The super-

conductor is patterned with a pinning potential $U(x, y) = U(x)$ which is periodic with period l along the x direction, has an asymmetric shape within one period, and is translationally invariant along the y direction of the sample. The simplest example of an asymmetric periodic potential, obtained for example by varying the sample thickness, is the asymmetric sawtooth potential (Fig. 1b). In the presence of a current with density J flowing along the y axis the vortices move with the velocity

$$\mathbf{v} = (\mathbf{f}_L + \mathbf{f}_{vv} + \mathbf{f}_u)/\eta \quad (1)$$

where $\mathbf{f}_L = (J \times \hat{\mathbf{h}})\Phi_0 d/c$ is the Lorentz force moving the vortices transverse to the current, c is the velocity of light in vacuum, $\hat{\mathbf{h}}$ is the unit vector pointing in the direction of the external magnetic field $\mathbf{H}$, $\mathbf{f}_u = -(dU/dx)\hat{\mathbf{x}}$ is the force generated by the periodic potential, $\mathbf{f}_{vv}$ is the repulsive vortex-vortex interaction, $\Phi_0 = 2.07 \times 10^{-7}$ G cm² is the flux quantum, η is the viscous drag coefficient, and d is the length of the vortices (that is, the thickness of the sample).

When a direct current flows along the positive y direction, the Lorentz force moves the vortices along the positive x direction with velocity v_+ . Reversing the current reverses the direction of the vortex velocity, but its magnitude, $|v_-|$, due to the asymmetry of the potential, is different from v_+ . For the sawtooth potential shown in Fig. 1b, the vortex velocity is higher when the vortex is driven to the right, than when it is driven to the left ($v_+ > |v_-|$). As a consequence the application of an alternating current (which is the consecutive application of direct and reverse currents with density J) results in a net velocity $v = (v_+ + v_-)/2$ to the right in Fig. 1b. This net velocity induced by the combination of an asymmetric potential and an a.c. driving force is called ratchet velocity⁷⁻¹⁰. The ratchet velocity for the case of low vortex density (when vortex-vortex interactions are neglected) can be calculated analytically. Denoting the period of the alternating current by T , the ratchet velocity of the vortices in the $T \rightarrow \infty$ limit is given by the expression

$$v = \begin{cases} 0 & \text{if } f_L < f_1 \\ \frac{1}{2\eta} \frac{(f_1 + f_2)(f_L - f_1)}{f_1 + f_2 - f_1} & \text{if } f_1 < f_L < f_2 \\ \frac{1}{\eta} \frac{f_1 f_2 (f_2 - f_1)}{f_1^2 - (f_2 - f_1)^2} & \text{if } f_2 < f_L \end{cases} \quad (2)$$

where $f_1 = \Delta U/l_1$ and $f_2 = \Delta U/l_2$ are the magnitudes of the forces generated by the ratchet potential on the facets of length l_1 and l_2 , respectively (Fig. 1c), ΔU is the energy difference between the

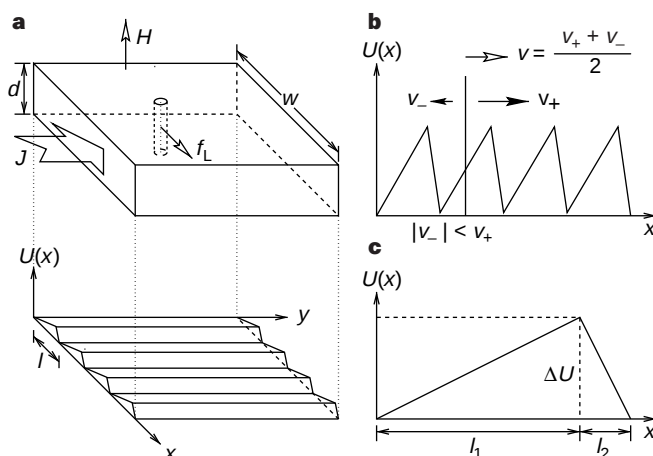


Figure 1 Patterning of a superconductor with an asymmetric potential. **a**, Diagram of a superconductor in the presence of an external magnetic field H . A direct current with density J flowing along the y direction (indicated by the large arrow) induces a Lorentz force f_L that moves the vortex in the x direction. The superconductor is patterned with a pinning potential $U(x, y) = U(x)$, whose shape is shown in the lower panel. The potential is periodic and asymmetric

along the x direction, and is translationally invariant along y . **b**, The pinning potential $U(x)$ along the superconductor's cross-section. The solid arrows indicate the vortex velocity v_+ (v_-) induced by a direct $+J$ (reversed $-J$) current. The average, $v = (v_+ + v_-)/2$, is the ratchet velocity of the vortex, obtained when an alternating current is applied. **c**, The parameters characterizing a single tooth of the asymmetric potential.

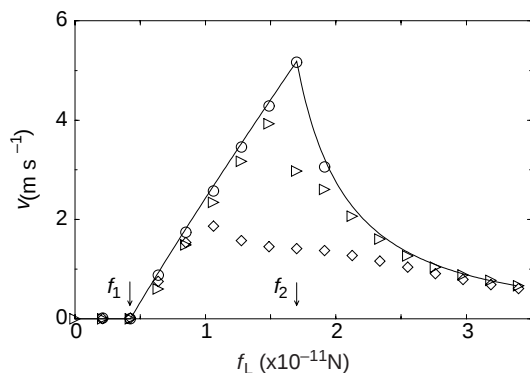


Figure 2 Ratchet velocity of the vortices as a function of the amplitude of the driving force f_L . The thick solid line corresponds to the analytical result (equation (2)) for a single vortex line. The symbols are the result of the numerical simulations for multiple vortices. The simulations were done using the model developed by Nori and collaborators^{20–22}, assuming that the rigid vortices are point-like objects moving in the x - y plane. At time zero the vortices are positioned randomly in the superconductor with a density ρ_0 and they move with velocity given by equation (1). The vortex-vortex interaction between two vortices at position $\mathbf{r}_i$ and $\mathbf{r}_j$ is modelled using $\mathbf{f}_{ij} = \Phi_0^2 d / (8\pi^2 \lambda^3 K_1) ((\mathbf{r}_i - \mathbf{r}_j) / \lambda) \hat{\mathbf{r}}_{ij}$, where $\hat{\mathbf{r}}_{ij} = (\mathbf{r}_i - \mathbf{r}_j) / |\mathbf{r}_i - \mathbf{r}_j|$. Here the modified Bessel function K_1 is cut off beyond the distance $r = 25\lambda$, where λ is the penetration depth (for Nb $\lambda = 45$ nm at $T = 0$). The force f generated by the sawtooth pinning potential shown in Fig. 1 is equal to f_1 when $kl < x < kl + l_1$, and f_2 when $kl + l_1 < x < (k+1)l$, where $k = 0, 1, \dots, N-1$. We choose $l_1 = 20\lambda = 0.9$ μm , $l_2 = 5\lambda = 0.225$ μm , $l = l_1 + l_2$ and $N = 10$, giving for the total width of the sample $w = 11.25$ μm . Its length (along the y direction) is set to 12 μm . The sample has periodic boundary conditions in both the x and y direction. The Lorentz force due to the alternating current is equal to $+f_L$ for $T/2$ time, and $-f_L$ for $T/2$ using $T = 0.3$ μs . We considered the simplest case, in which the potential is induced by thickness variations of a Nb superconductor thin film of thickness d , that is, the superconductor thickness, $d + h(x)$, changes along the x direction, following a sawtooth pattern. The pinning energy acting on the vortices is given by $U(x) = (d + h(x))\epsilon_0$, where ϵ_0 is the line energy of the vortex per unit length. Thus the magnitudes of the forces acting on the vortices are $f_1 = \epsilon_0 \Delta h / l_1$ and $f_2 = \epsilon_0 \Delta h / l_2$ for the two facets of the Δh high teeth (Fig. 1c), and we choose $\Delta h = l_2$. For Nb we have $\epsilon_0 = 1.7 \times 10^{-11}$ N, the viscosity per unit length is $\eta_0 = 7 \times 10^{-6}$ N s m^{-2} , yielding $\eta = \eta_0 d = 1.4 \times 10^{-12}$ N s m^{-1} for a $d = 2,000$ Å-thick film. The total number of vortices in the simulation were $n = 5$ (open circles), $n = 250$ (open triangles) and $n = 500$ (open diamonds) corresponding to a magnetic field in the sample of about 0.7 G, 35 G and 70 G, respectively.

maximum and the minimum of the potential, and $f_L = |f_L| = J\Phi_0 d/c$.

Because vortex-vortex interactions play an important role for high magnetic fields, we have performed molecular-dynamics simulations to determine the ratchet velocity for a collection of vortices. As Fig. 2 demonstrates, we find that for low vortex densities the numerical results follow closely the analytical prediction (equation (2)), and the magnitude of the ratchet velocity decreases with increasing vortex density. The vortex densities used in the simulations correspond to an internal magnetic field of about 0.7, 35 and 70 G, covering a wide range of magnetic fields. An important question for applications is if the ratchet velocity (equation (2)) is large enough to induce observable vortex motion at experimentally relevant timescales. To address this issue in Fig. 2 we plotted v for niobium, a typical low-temperature superconductor used in a wide range of devices, for which the potential $U(x)$ is induced by thickness variations of the superconductor. The details of the model are described in Fig. 2 legend. As the figure indicates, the maximum ratchet velocity (5.2 m s^{-1}) is high enough to move a vortex across the width of a typical sample (a few micrometres; ref. 3) in a few microseconds. Furthermore, increasing the vortex density by two orders of magnitude decreases the vortex velocity by only a factor of three.

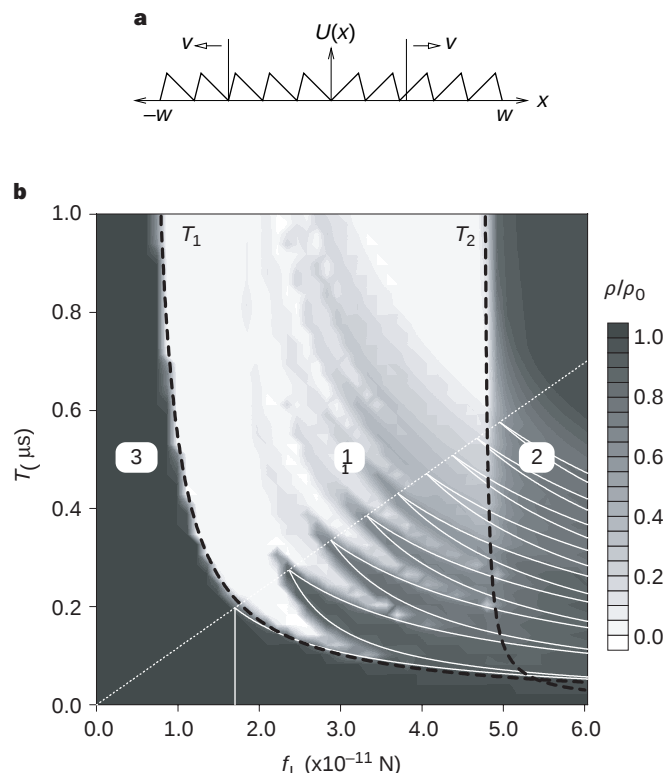


Figure 3 Removing vortices from a superconductor using an asymmetric potential. **a**, The potential necessary to remove vortices from the superconductor. Using the simulation method described in Fig. 2 legend, we investigated a system consisting of $N = 5$ teeth oriented to the left and the same number oriented to the right, as shown in the figure, the parameters of each tooth being identical to that described in Fig. 1c. To mimic the pressure generated by the external magnetic field, which acts to push vortices into the sample, on the two sides we attached two reservoirs, which have a constant vortex density ρ_0 at all times. Thus, vortices can leave the sample for the reservoir, or new vortices can enter from the reservoir. In thin superconducting films, due to the Meissner current, there is a geometrical barrier that acts to trap the vortices inside the sample²³. As most applications of superconductors involve thin films, we included in the simulations this geometrical barrier, which creates a force $f_{in}(x) = -(H\Phi_0/2\pi)\lambda/\sqrt{w^2 - x^2}$ for $-w + d/2 < x < w - d/2$, and $f_{edge} = 2\epsilon_0 - (H\Phi_0/2\pi)\sqrt{4wd/2 - 1}$ for $x > w - d/2$, and $-f_{edge}$ for $x < -w + d/2$. Thus the geometrical barrier opposes the entry of the vortices at the edge of the superconductor, but once they move inside, it moves them towards the centre of the superconductor. For successful vortex removal, the ratchet effect has to be strong enough to move the vortices against $f_{in}(x)$. **b**, The (f_L, T) diagram describing the effectiveness of the ratchet effect as a function of the parameters characterizing the driving current, f_L . The grey scale corresponds to the relative vortex density ρ/ρ_0 , where ρ_0 is the initial vortex density corresponding to $H = 1$ G and ρ is the final vortex density after the application of the alternating current. As the grey shading indicates, there is a region where vortex removal is complete, the vortex density being equal to zero. The dashed black lines correspond to the T_1 and T_2 boundaries, which are calculated analytically (see text) and separate the three main regimes: regime 1, complete vortex removal in much of the regime, $\rho = 0$; regime 2, partial vortex removal, $0 \leq \rho < \rho_0$; and regime 3, no change in the vortex density, $\rho = \rho_0$. The thin solid white lines denote the boundaries of the regions where vortex trapping occurs due to periodic orbits. These boundaries correctly reflect the structure of the fingers, but slightly deviate from the results of the numerical simulation, because the analytical calculation assumed an array of identical teeth.

Next, we discuss a potentially useful application of the ratchet effect by demonstrating that it could be used to drive vortices out of a superconductor. We consider a superconducting film that is patterned with two arrays of the ratchet potential oriented in opposite directions (Fig. 3a). During the application of the alternating current, the asymmetry of the potential in the right half

moves the vortices in that region to the right, while vortices in the left half move to the left. Thus the vortices drift towards the closest edge of the sample, decreasing the vortex density in the bulk of the film. We performed numerical simulations to quantitatively characterize this effect (details and the parameters are described in Fig. 3 legend). In Fig. 3b we summarize the effectiveness of vortex removal by plotting the reduced vortex density inside the film as a function of the Lorentz force f_L and the period T of the current. There is a well defined region where the vortex density drops to zero inside the sample, indicating that the vortices are completely removed from the bulk of the sample. Outside this region, we observe either a partial removal of the vortices or the alternating current has no effect on the vortex density.

The (f_L, T) diagram shown in Fig. 3b has three main regimes (1, 2 and 3) separated by two boundaries. The $T_1 = 2\eta l_1 / [f_L - [f_1 + f_{in}(-w + l_2)]]$ phase boundary (here we assume $d/2 < l_2$, w is the width of the sample, and $f_{in}(x)$ is defined in Fig. 3 legend) provides the time needed to move the vortex all the way up on the l_1 long facet of the ratchet potential at the edge of the superconductor, that is, to remove the vortex from the superconductor. When $T < T_1$, the vortices cannot leave the superconductor. The T_2 phase boundary defined as

$$T_2 = 2\eta \left(\frac{d/2}{f_L - [f_2 + f_{edge}]} + \frac{l_2 - d/2}{f_L - [f_2 - f_{in}(-w + d/2)]} \right)$$

where f_{edge} , also defined in Fig. 3, is the time needed for a vortex to enter from the edge of the superconductor past the first potential maxima. Thus, when $T < T_2$ the vortices cannot overcome the edge of the potential barrier. These phase boundaries, calculated for non-interacting vortices, effectively determine the vortex density in the three phases. Vortex removal is most effective in regime 1, where the vortices cannot move past the first potential barrier when they try to enter the superconductor, but they get past the barriers opposing their exit from it. Thus the vortices are swept out of the superconductor by the ratchet effect, and no vortex can re-enter, leading to a vortex density $\rho = 0$. Indeed, the numerical simulations indicate complete vortex removal in much of this phase (see the contour lines in Fig. 3b). An exception is the finger structure near the crossing of the T_1 and T_2 boundaries. For fields and periods within the first finger (situated almost entirely in regime 3 in Fig. 3b), the vortex follows a periodic orbit inside a single potential well¹⁰. The subsequent fingers represent stable period orbits between two, three or more wells, respectively. As the vortices cannot escape from these orbits, they remain trapped inside the superconductor, increasing the vortex density within the fingers in the phase diagram. Figure 3b shows (as white solid lines) the analytically calculated envelopes of the regions where such trapping occurs. An important feature of the finger structure is that stable periodic orbits, which prevent vortex removal, do not exist above the line $T_{tip} = f_L 2\eta \Delta U / [f_1 f_2 (f_2 - f_1)]$ connecting the finger tips. In regime 2 vortices can enter the superconductor, but the ratchet effect is still sweeping them out; so here we expect partial removal of the vortices, the final vortex density inside the superconductor being determined by the balance of vortex nucleation rate at the edge of the sample (which depends on its surface properties) and the ratchet velocity moving them out. In regime 3 the vortices cannot leave the superconductor and new vortices cannot enter the system, thus the initial density inside the superconductor is unchanged throughout this phase ($\rho = \rho_0$).

As the forces $f_{in}(x)$ and f_{edge} depend on H , the positions of the phase boundaries T_1 and T_2 also depend on the external magnetic field. In particular, there exists a critical field H^* for complete vortex removal. For $H > H^*$, regime 1 (where vortex removal is complete) disappears, but regime 2 with partial vortex removal survives. We find that for niobium films of the geometry described in Fig. 3 we have $H^* \approx 10$ G. But as H^* is a consequence of the geometric barrier, its value can be modified by changing the aspect ratio of the film.

Furthermore, for superconductors with elliptic cross-section the geometric barrier can be eliminated¹¹, thus phase 1 with complete vortex removal could be extended to high magnetic fields as well.

Vortex removal is important for numerous applications of superconductors and can improve the functioning of several devices. An immediate application of the proposed method would be improving the operation of superconducting quantum interference devices (SQUIDs), used as sensors in a wide assortment of scientific instruments^{3,12,13}. A long-standing issue in the performance of SQUIDs is $1/f$ noise^{6,12}, arising from the activated hopping of trapped vortices¹. Reducing the vortex density in these superconductors is expected to extend the operation regime of these devices to lower frequencies.

Although over the past few years several applications of the ratchet effect have been proposed, such as separating particles^{14,15}, designing molecular motors¹⁶, smoothing surfaces¹⁷, or rectifying voltage in Josephson junctions^{18,19}, our proposal solves an acute problem of condensed-matter physics, by removing vortices from a superconductor. In contrast with most previous applications, which require the presence of thermal noise, our model is completely deterministic. Indeed, in niobium the variation in the pinning potential is $\Delta U \approx 25$ eV, which is more than 10^4 times larger than $k_B T \approx 0.8$ meV at $T_c = 9.26$ K, thus rendering thermal fluctuations irrelevant. A particularly attractive practical feature of our proposed method is that it does not require sophisticated material processing to make it work. First, it requires standard scale (micrometre scale) patterning techniques; the micrometre tooth size was chosen so that a few teeth fit on a typical SQUID, but larger feature size will also function if the period T is increased proportionally). Second, the application of an alternating current with appropriate period and intensity is rather easy to achieve. For applications where an alternating current is not desired, the vortices can be 'flushed out' before the normal operation of the device. On the other hand, if the superconducting device is driven by an alternating current (for example, radio-frequency SQUIDs, a.c. magnets, or wires carrying alternating current), the elimination of the vortices will take place continuously during the operation of the device. The analytically predicted phase boundaries, whose position is determined by the geometry of the patterning, provide a useful tool for designing the appropriate patterning to obtain the lowest possible vortex density for current and frequency ranges desired for specific applications. Finally, although here we limited ourselves to low-temperature superconductors, the working principle of the ratchet effect also applies to high-temperature superconductors. □

Received 22 December 1998; accepted 10 May 1999.

1. Tinkham, M. *Introduction to Superconductivity* (McGraw-Hill, New York, 1996).
2. Donaldson, G. B., Pegrum, C. M. & Bain, R. J. P. in *SQUID '85: Proc. 3rd Int. Conf. on Superconducting Quantum Devices* (eds Hahlbohm, H. D. & Lübbig, H.) 749–753 (Walter Gruyter, Berlin, 1985).
3. Clarke, J. *Superconducting Devices* (eds Ruggiero, S. T. & Rudman, D. A.) 51–100 (Academic, 1990).
4. Blatter, G., Feigelman, M. V., Geshkenbeim, V. B., Larkin, A. I. & Vinokur, V. M. Vortices in high temperature superconductors. *Rev. Mod. Phys.* **66**, 1125–1388 (1994).
5. Koch, R. H., Sun, J. Z., Foglietta, V. & Gallagher, W. J. Flux dam, a method to reduce extra low-frequency noise when a superconducting magnetometer is exposed to a magnetic field. *Appl. Phys. Lett.* **67**, 709–711 (1995).
6. Dantsker, E., Tanaka, S. & Clarke, J. High- T_c superconducting quantum interference devices with slots or holes: low $1/f$ noise in ambient magnetic fields. *Appl. Phys. Lett.* **70**, 2037–2039 (1997).
7. Magnasco, M. O. Forced thermal ratchet. *Phys. Rev. Lett.* **71**, 1477–1481 (1993).
8. Astumian, R. D. Thermodynamics and kinetics of a brownian motor. *Science* **276**, 917–922 (1997).
9. Jülicher, F., Ajdari, A. & Prost, J. Modeling molecular motors. *Rev. Mod. Phys.* **69**, 1269–1281 (1997).
10. Hänggi, P. & Bartsussek, R. in *Lecture Notes in Physics* Vol. 476 (eds Parisi, J. et al.) 294–308 (Springer, Berlin, 1996).
11. Clem, J. R., Huebener, R. P. & Gallus, D. E. Gibbs free-energy barrier against irreversible magnetic flux entry into a superconductor. *J. Low Temp. Phys.* **12**, 449–477 (1973).
12. Mück, M. Practical aspects for SQUID applications. *Superlattices Microstruct.* **21**, 415–421 (1997).
13. Scott, B. A., Kirtley, J. R., Walker, D., Chen, B.-H. & Wang, Y. Application of scanning SQUID petrology to high-pressure materials science. *Nature* **389**, 164–167 (1997).
14. Rousselet, J., Salome, L., Ajdari, A. & Prost, J. Directional motion of brownian particles induced by a periodic asymmetric potential. *Nature* **370**, 446–448 (1994).
15. Faucheux, L. P., Bourdieu, L. S., Kaplan, P. D. & Libchaber, A. J. Optical thermal ratchet. *Phys. Rev. Lett.* **74**, 1504–1507 (1995).
16. Kelly, T. R., Silva, H. D. & Silva, R. A. A rationally designed molecular motor. *Nature* (submitted).
17. Derényi, I., Lee, C.-S. & Barabási, A.-L. Ratchet effect in surface electromigration: smoothing surfaces by an AC field. *Phys. Rev. Lett.* **80**, 1473–1476 (1998).
18. Zapata, I., Bartsussek, R., Sols, F. & Hänggi, P. Voltage rectification by a SQUID ratchet. *Phys. Rev. Lett.* **77**, 2292–2295 (1996).

19. Zapata, I., Luczka, J., Sols, F. & Hänggi, P. Tunneling center as a source of voltage rectification in Josephson junctions. *Phys. Rev. Lett.* **80**, 829–832 (1998).
20. Reichhardt, C., Olson, C. J. & Nori, F. Nonequilibrium dynamic phases and plastic flow of driven vortex lattices in superconductors with periodic arrays of pinning sites. *Phys. Rev. B* **58**, 6534–6564 (1998).
21. Reichhardt, C., Olson, C. J. & Nori, F. Dynamic phases of vortices in superconductors with periodic pinning. *Phys. Rev. Lett.* **78**, 2648–2651 (1997).
22. Olson, C. J., Reichhardt, C. & Nori, F. Nonequilibrium dynamic phase diagram for vortex lattices. *Phys. Rev. Lett.* **81**, 3757–3760 (1998).
23. Zeldov, E. *et al.* Geometrical barriers in high-temperature superconductors. *Phys. Rev. Lett.* **73**, 1428–1431 (1994).

Acknowledgements. We thank R. D. Astumian, D. J. Bishop, S. N. Coppersmith, D. Grier, H. Jeong, A. Koshelev and S. T. Ruggiero for discussions and help during the preparation of the manuscript. This research was supported by an NSF Career Award.

Correspondence and requests for materials should be addressed to A.-L.B. (e-mail: alb@nd.edu).

Synthesis of cubic silicon nitride

Andreas Zerr*, Gerhard Miehe†, George Serghiou‡, Marcus Schwarz*, Edwin Kroke*, Ralf Riedel*, Hartmut Fueß†, Peter Kroll§ & Reinhard Boehler‡

* Fachgebiet Disperse Feststoffe, † Fachgebiet Strukturforchung, Fachbereich Materialwissenschaft, Technische Universität Darmstadt, Petersenstrasse 23, D-64287 Darmstadt, Germany

‡ Max-Planck-Institut für Chemie, Postfach 3060, D-55020 Mainz, Germany

§ Cornell University, Department of Chemistry, Baker Laboratory, Ithaca, New York 14853-1301, USA

Silicon nitride (Si_3N_4) is used in a variety of important technological applications. The high fracture toughness, hardness and wear resistance of Si_3N_4 -based ceramics are exploited in cutting tools and anti-friction bearings¹; in electronic applications, Si_3N_4 is used as an insulating, masking and passivating material². Two polymorphs of silicon nitride are known, both of hexagonal structure: α - and β - Si_3N_4 . Here we report the synthesis of a third polymorph of silicon nitride, which has a cubic spinel structure. This new phase, $c\text{-Si}_3\text{N}_4$, is formed at pressures above 15 GPa and temperatures exceeding 2,000 K, yet persists metastably in air at ambient pressure to at least 700 K. First-principles calculations of the properties of this phase suggest that the hardness of $c\text{-Si}_3\text{N}_4$ should be comparable to that of the hardest known oxide (stishovite³, a high-pressure phase of SiO_2), and significantly greater than the hardness of the two hexagonal polymorphs.

The spinel structure is found in many binary and ternary metal oxides, silicates and sulphides, such as Fe_3O_4 , $\gamma\text{-Mg}_2\text{SiO}_4$ or Fe_3Si_4 (ref. 4). A mixed oxide/nitride spinel phase is formed from Al_2O_3 and AlN which is known as ALON ($\text{Al}_{23}\text{O}_{27}\text{N}_5$; ref. 5). The highest proportion of nitrogen anions (up to 75%) is contained in another oxide/nitride spinel, $\text{Mn}_2(\text{MnTa}_3)\text{N}_{6-\delta}\text{O}_{2-\delta}$ ($0 \leq \delta \leq 1$)⁶. In oxynitride spinels, as well as in other structures such as the solid solution of Si_3N_4 and Al_2O_3 denoted as $\beta\text{-SiALON}$, the silicon atoms are coordinated to both nitrogen and oxygen. The only oxide/nitride compound containing pure SiN_6 octahedra is, to our knowledge, the nitrido silicate $\text{Ce}_{16}\text{Si}_{15}\text{O}_6\text{N}_{32}$, which crystallizes in a perovskite-like structure⁷. However, pure nitride-based, non-oxide solid silicon compounds containing six-fold coordinated silicon atoms have not so far been reported.

Particular interest in the high-pressure phase behaviour of Si_3N_4 arises from a theoretical prediction⁸ that the hardness of hypothetical $\beta\text{-C}_3\text{N}_4$ (with the hexagonal $\beta\text{-Si}_3\text{N}_4$ structure) may be comparable to that of diamond. It was suggested in a more recent theoretical study⁹ that cubic C_3N_4 with the structure of the high-pressure phase of Zn_2SiO_4 (willemite II, space group $I43m$) has a bulk modulus (and possibly a hardness) greater than that of diamond.

The compound $\alpha\text{-Si}_3\text{N}_4$, with a density of $\rho = 3.183 \text{ g cm}^{-3}$ (ref. 10) has been synthesized at ambient pressure and temperatures

below 1,800 K, while the β -phase ($\rho = 3.200 \text{ g cm}^{-3}$ (ref. 10)) requires higher temperatures^{11,12}. The relative stability of these phases and the influence of oxygen on the stability of $\alpha\text{-Si}_3\text{N}_4$ is not well understood. This subject is discussed in more detail in ref. 1. Both α - and $\beta\text{-Si}_3\text{N}_4$ maintain their structure during compression at room temperature up to 48 GPa (ref. 13) and 28 GPa (A.Z., R.B. and R.R., manuscript in preparation), respectively. The β -phase has been synthesized at 1.5 GPa and 1,400 K from silicon and nitrogen using YAG-laser heating in a diamond cell¹⁴, and from $\alpha\text{-Si}_3\text{N}_4$ at ~ 5 GPa and 2,270 K using an internally heated high-pressure vessel¹². No high-temperature experiments on silicon nitride above 9 GPa have yet been performed.

To investigate the high-pressure and high-temperature phase behaviour of silicon nitride, we utilized the technique of laser heating in a diamond cell. In all experiments, silicon single crystals or cold-pressed powder pellets of amorphous Si_3N_4 and polycrystalline α - and $\beta\text{-Si}_3\text{N}_4$ were placed in a nitrogen pressure medium. Nitrogen was used for two reasons: first, for the case where silicon was used as a starting material, it served as a reactant for the synthesis of silicon nitride. Second, it precludes decomposition of silicon nitride starting materials on heating. A neodymium:yttrium-lithium-fluoride laser (Nd:YLF; Antares-YLF, Coherent, Palo Alto) was used for heating silicon, and a CO_2 laser (Melles Griot, Carlsbad, California) for heating silicon nitride. Details of the laser heating techniques, temperature measurement, and sample pressure–temperature conditions are given elsewhere^{15,16}. The starting crystalline materials were characterized by Raman spectroscopy^{11,17}. The samples were heated at constant pressure for 1–10 min and then quenched by switching off the laser power. The reaction products were examined by Raman spectroscopy at high pressure and on pressure decrease. The recovered samples were investigated using a transmission electron microscope (TEM; Philips CM20UT) equipped with an energy dispersive X-ray (EDX) detector (HPGe, Voyager, Noran Instruments, Middleton, Wisconsin). TEM and EDX analysis allowed us to determine the crystal structure, and to obtain semiquantitative information on the chemical composition.

In the first experiment, elemental silicon was heated at 5.2 GPa using the Nd:YLF laser. Above 2,100 K, temperature fluctuations indicated a chemical reaction of silicon with the nitrogen pressure

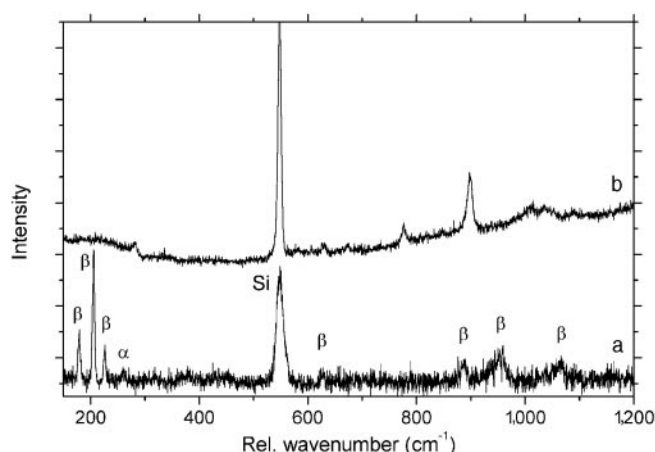


Figure 1 High-pressure Raman spectra of silicon nitride phases synthesized by heating elemental silicon and molecular nitrogen in a diamond cell. Trace a; At 5.2 GPa we found, in addition to the Raman band of silicon (labelled 'Si'), seven new lines corresponding to hexagonal $\beta\text{-Si}_3\text{N}_4$ and one new weak line that can be assigned to the most intense Raman line of hexagonal $\alpha\text{-Si}_3\text{N}_4$ at this pressure. Trace b; after heating at 15 GPa none of the Raman lines obtained can be attributed to the known phases of silicon or α - or $\beta\text{-Si}_3\text{N}_4$, indicating the appearance of a new cubic silicon nitride phase with spinel-type structure (see text and Fig. 2).

medium. Examination of the reaction product with an optical microscope showed, instead of the starting opaque silicon crystal, a pale grey powder. High-pressure Raman spectra of the reaction product indicated a mixture of β - Si_3N_4 and possibly a small amount of α - Si_3N_4 (ref. 11, and A.Z., R.B. and R.R., manuscript in preparation) in addition to the starting silicon (Fig. 1, trace a). The second Nd:YLF laser heating experiment with silicon at 15 GPa and 2,200 K—which is $\sim 1,000$ K above the melting temperature of elemental silicon at that pressure¹⁸—yielded a light-yellow transparent sphere 15 μm in diameter as the reaction product. The Raman spectrum of this material (Fig. 1, trace b) could not be assigned to either α - or β - Si_3N_4 . Furthermore, no peaks indicating the presence of elemental silicon were found either at 15 GPa (at which pressure metallic Si II exhibits Raman bands at ~ 110 and 410 cm^{-1} ; ref. 19) or after decompression (when Si II would be expected to transform back to the body-centred cubic phase Si III, with Raman lines between 300 and 500 cm^{-1} ; refs 17, 20).

The recovered reaction product was ground between two diamond anvils for TEM analysis which showed a new cubic silicon nitride phase for all grains examined. EDX analyses of the grains revealed only the presence of silicon and nitrogen with a Si:N ratio identical to that of β - Si_3N_4 . From selected-area electron diffraction patterns, a cubic face-centred cell with a lattice constant of $a_0 = 7.80 \pm 0.08\text{ \AA}$ was derived. The [001] zone exhibits extinctions due to a d -glide plane (Fig. 2). Accordingly, the possible space groups of cubic Si_3N_4 are $Fd\bar{3}$, $Fd\bar{3}m$ or $Fd\bar{3}c$. We exclude the space group $Fd\bar{3}c$ because all sites with multiplicities < 96 lead to the extra condition that for all possible reflections, h, k, l must be even. This contradicts our observation of the first-order Laue zone (h, k, l) (Fig. 2). Hence a stoichiometry of Si_3N_4 can be realised in $Fd\bar{3}c$ only with a number of formula units per unit cell $Z \geq 24$; the corresponding density would exceed 11.8 g cm^{-3} , which is unrealistic.

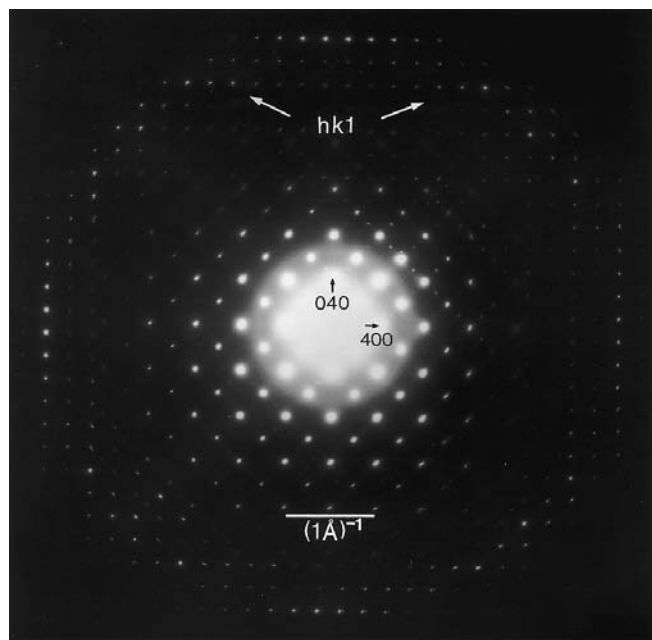


Figure 2 Selected-area electron diffraction pattern of zone [001] of the new cubic silicon nitride. Observation of the first-order Laue zone ($hk1$) and extinctions due to the d -glide plane require the structure to have the space group $Fd\bar{3}m$ with $Z = 8$. The (040) and (400) reflections are indicated by arrows. The positions of the Si and N atoms can be described in two different but equivalent ways. (1) Si(1): $8a(1/8, 1/8, 1/8)$, Si(2): $16d(1/2, 1/2, 1/2)$, N: $32e(x, x, x)$ with $x = 0.25 + \delta$. (2) Si(1): $8b(3/8, 3/8, 3/8)$, Si(2): $16c(0, 0, 0)$, N: $32e(x, x, x)$ with $x = 0.25 - \delta$. This is the spinel structure. The intensities needed to determine the small deviation δ of $x(\text{N})$ from 0.25 by structure refinement are not available. Nevertheless, as for $\delta = 0.0125$ the Si–N distances for the tetrahedrally and the octahedrally coordinated Si atoms became equal, δ must be smaller than this value.

In the space group $Fd\bar{3}m$ and in its subgroup $Fd\bar{3}$, however, there are two equivalent ways to describe a realistic structure for the Si_3N_4 stoichiometry with $Z = 8$, which both lead to the spinel structure (Fig. 2). For $c\text{-Si}_3\text{N}_4$ with spinel structure the calculated density is $3.93 \pm 0.12\text{ g cm}^{-3}$, which is 23% higher than that of α - or β - Si_3N_4 . Comparison of the high-resolution TEM images with results of image simulation (not shown) confirmed the spinel structure for $c\text{-Si}_3\text{N}_4$. Because the silicon atoms in the spinel phase are four- and six-fold coordinated by nitrogen in a 1:2 ratio, the cubic form of silicon nitride may be better described as $\text{Si}[\text{Si}_2\text{N}_4]$ rather than Si_3N_4 .

We also performed CO_2 -laser heating experiments with amorphous Si_3N_4 , as well as α - and β - Si_3N_4 , at 15 and 30 GPa and up to 2,800 K. For all these materials, the Raman spectra of the heated samples showed in addition to the bands of the starting phases (α - or β - Si_3N_4) a new band at around 548 cm^{-1} which is the strongest band of $c\text{-Si}_3\text{N}_4$ (Fig. 1, trace b). Furthermore, synthesis of $c\text{-Si}_3\text{N}_4$ at 30 GPa was verified by TEM analysis of the recovered sample. Thus $c\text{-Si}_3\text{N}_4$ is stable at high temperatures and at pressures between 15 and 30 GPa. However, the lowest pressure at which $c\text{-Si}_3\text{N}_4$ can be synthesized has not yet been identified. We also examined the thermal metastability of $c\text{-Si}_3\text{N}_4$ in air at ambient pressure. Samples synthesized in a diamond cell (about $50 \times 10^{-9}\text{ g}$ in weight) were heated in an oven in air, and analysed subsequently by Raman spectroscopy. Our preliminary results revealed metastability of $c\text{-Si}_3\text{N}_4$ to at least 700 K, but the upper temperature limit for its metastability has not yet been determined.

We performed first-principles calculations to understand why cubic Si_3N_4 has the spinel structure and not the willemite II structure as predicted for cubic- C_3N_4 . (We note that the spinel structure has not been considered theoretically for C_3N_4 in the literature so far.) Our calculations enabled us to estimate the compressibility and the shear c_{44} modulus of $c\text{-Si}_3\text{N}_4$, which are related to its hardness. The calculations were based on the method described in ref. 9, and yield the dependence of the total energy on volume or on shear deformation of the lattice, from which we derived bulk and single crystal moduli for a chosen structure^{21,22}.

The calculated energy-versus-volume (E - V) curves for three silicon nitride modifications— β - Si_3N_4 , willemite II ($w\text{-Si}_3\text{N}_4$) and spinel structure ($c\text{-Si}_3\text{N}_4$)—are shown in Fig. 3 over a wide range of volume per formula unit. As $\alpha\text{-Si}_3\text{N}_4$ and $\beta\text{-Si}_3\text{N}_4$ are very similar in their E - V diagram, only that of the β -phase is given in Fig. 3. Comparing the E - V curves calculated for $w\text{-Si}_3\text{N}_4$ and $c\text{-Si}_3\text{N}_4$, it is evident that at high pressures the latter is thermodynamically stable. A common tangent construction²¹ of the E - V curves gives the pressures p_t for the phase transitions $\beta\text{-Si}_3\text{N}_4 \rightarrow c\text{-Si}_3\text{N}_4$ and $\beta\text{-Si}_3\text{N}_4 \rightarrow w\text{-Si}_3\text{N}_4$ to be 13 and 20 GPa,

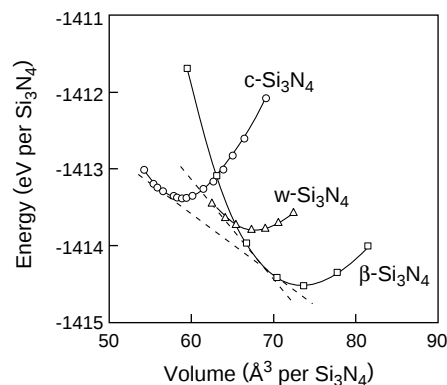


Figure 3 Calculated energy-versus-volume data for three Si_3N_4 structures. Squares represent results for hexagonal $\beta\text{-Si}_3\text{N}_4$, circles for $c\text{-Si}_3\text{N}_4$, and triangles for hypothetical $w\text{-Si}_3\text{N}_4$. The solid lines are fits of the Murnaghan equation of state²³ to the E - V data. The dashed lines are common tangents which allow us to derive the phase transition pressures p_t (see text).

respectively. Thus on compression, β - Si_3N_4 will directly transform to α - Si_3N_4 and not to the hypothetical w - Si_3N_4 with lower density. This finding agrees with our synthesis of α - Si_3N_4 above 15 GPa.

The structural parameters for the optimized geometry of α - Si_3N_4 are $a = 7.76 \text{ \AA}$ and $\delta = 0.0074$. Fitting a Murnaghan equation of state²³ to the calculated $E-V$ data we obtained a bulk modulus of 300 GPa. This is about 20–30% higher than the experimental values reported for α - Si_3N_4 (229 GPa; ref. 13) and β - Si_3N_4 (250 GPa; ref. 24). Our calculations of the bulk moduli of the α - and β -phase gave 227 GPa and 249 GPa, respectively. Furthermore, we calculated the shear modulus c_{44} to be 340 GPa for α - Si_3N_4 and 150 GPa for β - Si_3N_4 .

The calculated bulk and shear c_{44} moduli for α - Si_3N_4 are close to those of stishovite, a high-pressure phase of SiO_2 where the silicon atoms are octahedrally coordinated to oxygen. The high coordination leads to a significant increase in density (4.29 g cm^{-3} ; ref. 4), bulk modulus (281–313 GPa; ref. 25) and c_{44} modulus (252 GPa; ref. 26) with respect to the low-pressure modifications of SiO_2 (such as quartz). Moreover, it was found recently that the hardness of SiO_2 -stishovite exceeds that of any other known oxide (Knoop hardness 33 GPa; ref. 3). Because the compressibility and the c_{44} modulus of α - Si_3N_4 and SiO_2 -stishovite are comparable, we expect that the hardness of α - Si_3N_4 will be close to that of stishovite, which is considered to be the third hardest material after diamond and cubic BN (ref. 3). Because of this, and because of its metastability in air at ambient pressure and high temperatures, α - Si_3N_4 can be considered for technological applications. □

Methods

We calculated²⁷ the total energies within the local-density approximation (LDA) and the generalized gradient approximation (GGA) of the exchange and correlation energy with standard plane-wave ($E_{\text{cutoff}} = 70 \text{ Ry}$) pseudopotential techniques. Special k -points (10 for α - Si_3N_4 and w - Si_3N_4 , 15 for β - Si_3N_4) were used to integrate over the Brillouin zone. At a given volume, the atomic positions were relaxed and in addition for β - Si_3N_4 the c/a ratio was optimized. A detailed description of the method, further calculations and discussions, including differences in results between the LDA or GGA, will be presented elsewhere.

Received 19 February; accepted 25 May 1999.

- Komeya, K. & Matsui, M. in *Materials Science and Technology* (eds Cahn, R. W., Haasen, P. & Kramer, E. J.) 518–565 (Wiley-VCH, Weinheim, 1994).
- Gmelin *Handbook of Inorganic and Organometallic Chemistry* Si Suppl. B 5c, *Silicon Nitride in Electronics* (Springer, Berlin, 1991).
- Leger, J. M. et al. Discovery of hardest known oxide. *Nature* **383**, 401 (1996).
- Smyth, J. R. & McCormick, T. C. in *Mineral Physics and Crystallography: a Handbook of Physical Constants* (ed. Ahrens, T. J.) 1–17 (Am. Geophys. Union, Washington DC, 1995).
- McCauley, J. W. & Corbin, N. D. Phase relations and reaction sintering of transparent cubic aluminum oxynitride spinel (ALON). *J. Am. Ceram. Soc.* **62**, 476–479 (1979).
- Grins, J., Käll, P.-O. & Svensson, G. Synthesis, structure and magnetic susceptibility of the oxynitride spinel $\text{Mn}_2(\text{MnTa}_3)\text{N}_{6-3}\text{O}_{2+3}$, $0 < \delta < 1$. *J. Solid State Chem.* **117**, 48–54 (1995).
- Köllisch, K. & Schnick, W. $\text{Ce}_{10}\text{Si}_5\text{O}_6\text{M}_{32}$ -An oxonitridosilicate with silicon octahedrally coordinated by nitrogen. *Angew. Chem. Int. Edn* **38**, 357–359 (1999).
- Liu, A. Y. & Cohen, M. L. Prediction of new low compressibility solids. *Science* **245**, 841–842 (1989).
- Teter, D. M. & Hemley, R. J. Low-compressibility carbon nitrides. *Science* **271**, 53–55 (1996).
- Powder Diffraction File-2 Database* (JCPDS Int. Centre for Diffraction Data, Newtown Square, PA 19073, USA, 1996).
- Wada, N., Solin, S. A., Wong, J. & Prochazka, S. Raman and IR absorption spectroscopic studies on α , β and amorphous Si_3N_4 . *J. Non-Cryst. Solids* **43**, 7–15 (1981).
- Bradley, R. S., Munro, D. C., Whitfield, M. The reactivity and polymorphism of selected nitrides at high temperatures and high pressures. *J. Inorg. Nucl. Chem.* **28**, 1803–1812 (1966).
- Kruger, M. B. et al. Equation of state of α - Si_3N_4 . *Phys. Rev. B* **55**, 3456–3460 (1997).
- Yoo, C.-S., Akella, J. & Nicol, M. in *Proc. 3rd Int. Conf. on Advanced Materials* (eds Akaishi, M. et al.) 175–179 (NIRIM, Tsukuba, Japan, 1996).
- Boehler, R., von Bargen, N. & Chopelas, A. Melting, thermal expansion and phase transitions of iron at high pressures. *J. Geophys. Res.* **95**, 21731–21736 (1990).
- Boehler, R. & Chopelas, A. A new approach to laser heating in high pressure mineral physics. *Geophys. Res. Lett.* **18**, 1147–1150 (1991).
- Mernagh, T. P. & Liu, L.-G. Pressure dependence of Raman phonons of some group IVA (C, Si, and Ge) elements. *J. Phys. Chem. Solids* **52**, 507–512 (1991).
- Bundy, F. P. Phase diagrams of silicon and germanium to 200 kbar, 1000 °C. *J. Chem. Phys.* **41**, 3809–3814 (1964).
- Olijnyk, H. Raman scattering in metallic Si and Ge up to 50 GPa. *Phys. Rev. Lett.* **68**, 2232–2234 (1992).
- Kobliska, R. J. et al. Raman scattering from phonons in polymorphs of Si and Ge. *Phys. Rev. Lett.* **29**, 725–728 (1972).
- Yin, M. T. & Cohen, M. L. Theory of static structural properties, crystal stability, and phase transformations: Application to Si and Ge. *Phys. Rev. B* **26**, 5668–5687 (1982).
- Ihm, J. Total energy calculations in solid state physics. *Rep. Prog. Phys.* **51**, 105–142 (1988).
- Murnaghan, F. D. The compressibility of media under extreme pressures. *Proc. Natl Acad. Sci. USA* **30**, 244–247 (1944).

- Cartz, L. & Jorgensen, J. D. The high-pressure behaviour of α -quartz, oxynitride, and nitride structures. *J. Appl. Phys.* **52**, 236–244 (1981).
- Knittle, E. in *Mineral Physics and Crystallography: a Handbook of Physical Constants* (ed. Ahrens, T. J.) 98–142 (Am. Geophys. Union, Washington DC, 1995).
- Bass, J. D. in *Mineral Physics and Crystallography: a Handbook of Physical Constants* (ed. Ahrens, T. J.) 45–63 (Am. Geophys. Union, Washington DC, 1995).
- Bockstedte, M., Kley, A., Neugebauer, J. & Scheffler, M. Density-functional theory calculations for poly-atomic systems: electronic structure, static and elastic properties and ab-initio molecular dynamics. *Comput. Phys. Commun.* **107**, 187–222 (1997).

Acknowledgements. We thank O. Tschauer and I.-W. Chen for discussions, and Bayer AG for providing the β - Si_3N_4 powder. This work was supported by the Deutsche Forschungsgemeinschaft, Bonn, Germany, and the Fonds der Chemischen Industrie, Frankfurt, Germany.

Correspondence and requests for materials should be addressed to R.R. (e-mail: dg9b@hrzpub.tu-darmstadt.de).

Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens

Jianwei Miao*, Pambos Charalambous†, Janos Kirz* & David Sayre*‡

* Department of Physics and Astronomy, State University of New York, Stony Brook, New York 11794-3800, USA

† Kings College, Strand, London WC2R 2LS, UK

The contrast and penetrating power afforded by soft X-rays when they interact with matter makes this form of radiation ideal for studying micrometre-sized objects^{1,2}. But although soft X-rays are useful for probing detail too fine for visible light microscopy in specimens too thick for electron microscopy, the highest-resolution applications of X-ray imaging have been traditionally limited to crystalline samples. Here we demonstrate imaging (at ~75 nm resolution) of a non-crystalline sample, consisting of an array of gold dots, by measuring the soft X-ray diffraction pattern from which an image can be reconstructed. The crystallographic phase problem³—the usually unavoidable loss of phase information in the diffraction intensity—is overcome by oversampling⁴ the diffraction pattern, and the image is obtained using an iterative algorithm⁵. Our X-ray microscopy technique requires no high-resolution X-ray optical elements or detectors. We believe that resolutions of 10–20 nm should be achievable; this would provide an imaging resolution about 100 times lower than that attainable with conventional X-ray crystallography, but our method is applicable to structures roughly 100 times larger. This latter feature may facilitate the imaging of small whole cells or large subcellular structures in cell biology.

When an object is illuminated by a plane wave, the amplitude of the far-field diffraction pattern is the Fourier transform of the object. The microscopist typically uses a lens to perform the inverse transform to create the image. The resolution of a perfect lens is limited by diffraction to $d > 0.61\lambda/\text{NA}$, where λ is the wavelength and NA is the numerical aperture of the lens. For the visible region of the spectrum, there are high-NA lenses which provide imaging near the limit set by the wavelength, and further improvements are possible using confocal and video techniques.

To get to considerably higher resolution, microscopists use shorter wavelengths, such as are obtained from electrons and X-rays. Transmission electron microscopes routinely image specimens up to ~0.5 μm thick, whereas X-ray microscopes are particularly useful for somewhat thicker specimens which may also be wet^{1,2,6,7}. The resolution of X-ray microscopes has been limited by the available optical elements ('optics') to about 30–50 nm (refs 8, 9), and for three-dimensional imaging to about 50–100 nm (refs 10–12) or,

‡ Present address: 15 Jefferson Court, Bridgewater, New Jersey 08807-3050, USA.

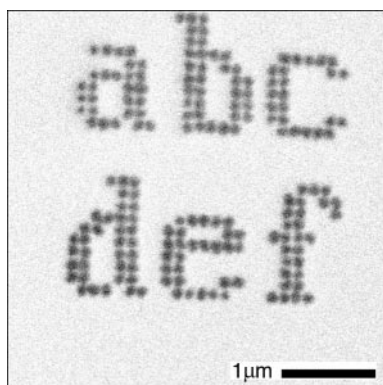


Figure 1 A scanning electron microscope image of the specimen. The specimen was fabricated by depositing gold dots, each ~ 100 nm in diameter and 80 nm thick, on a silicon nitride membrane.

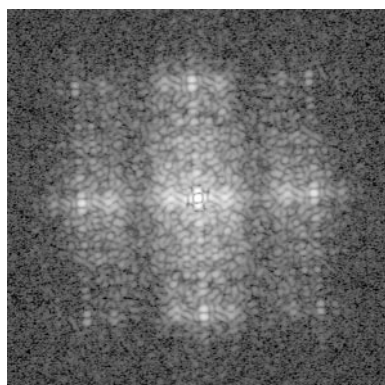


Figure 2 A diffraction pattern of the specimen (using a logarithmic intensity scale). The central 15-pixel-radius circular area is supplied by the squared magnitude of the Fourier transform of the optical microscope image (Fig. 3).

in Gabor holography, by the resolution of the detector and/or the optics^{13,14}.

X-ray crystallography is widely used for very-high-resolution imaging of molecular structure, when these molecules are arranged in a regular crystalline array. This technique requires no optics, and does not impose stringent resolution requirements on the detector. The multiple copies of the specimen amplify the signal, and their regular arrangement concentrates the far-field diffraction pattern into discrete Bragg peaks.

That the general approach of crystallography could be extended to image non-crystalline specimens was first proposed by Sayre in 1980¹⁵. This approach, although it poses particular challenges in recording and reconstructing the diffraction pattern¹⁶, has the potential to extend the resolution of three-dimensional X-ray microscopy beyond the technical limitations mentioned above, through extending the class of structures to which the powerful techniques of X-ray crystallography can be applied. Here we report the first (to our knowledge) successful recording and reconstruction of such an X-ray diffraction pattern.

The specimen was a collection of gold dots, each ~ 100 nm in diameter and 80 nm thick, deposited on a 100-nm-thick silicon nitride membrane, to form a set of six letters (Fig. 1). This specimen was illuminated with $\lambda = 1.7$ nm monochromatic and parallel X-rays from the X1A undulator beamline at the National Synchrotron Light Source. To ensure spatial coherence of the illumination, a 10- μ m-diameter pinhole was placed at a distance of 2.5 cm upstream of the specimen. To limit the effect of the scatter from the edge of this pinhole, the specimen was placed only ~ 30 μ m from the corners of the silicon nitride membrane, allowing the

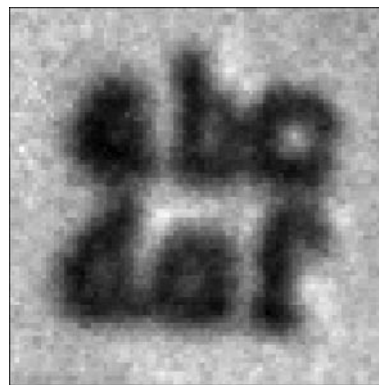


Figure 3 An optical microscope image of the specimen.

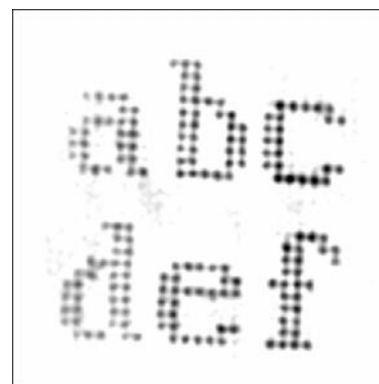


Figure 4 The specimen image as reconstructed from the diffraction pattern of Fig. 2.

silicon support to protect three quadrants of the detector from the scatter. In addition, the pinhole served to reduce the size of beam-stop needed to protect the detector from the direct beam; a 220- μ m-diameter wire placed in front of the detector was used as beam stop. The detector—a back-thinned, liquid-nitrogen-cooled CCD with 512×512 pixels and a $24 \mu\text{m} \times 24 \mu\text{m}$ pixel size—was placed downstream of the specimen at a distance of 25 cm. A photodiode could be inserted between the beam stop and the CCD to monitor the beam intensity in the presence or absence of pinhole, beam stop and specimen, and to help align these components in the beam. The apparatus was in vacuum with an air-lock for rapid sample change. A typical diffraction pattern is shown in Fig. 2, in which the fourth-quadrant data were obtained by using central symmetry. Since the central region was obscured by the beam stop, we replaced the central area of the X-ray diffraction pattern by a patch from the squared magnitude of the Fourier transform of an optical microscope image of the specimen shown in Fig. 3. The patch is a circular area with a 15-pixel radius, which occupies less than 0.5% of the whole diffraction pattern. The exposure time of the pattern was 15 min which corresponds to an incident flux of 1.65×10^{19} photons m^{-2} , or a radiation dose of 1.56×10^6 gray (Gy) to the specimen. The pattern extends to the edge of the CCD, suggesting that a larger detector would directly lead to higher spatial resolution.

The intensity of the diffraction pattern provides a record of the size, but not the phase, of the diffraction amplitude. To reconstruct the image, one faces therefore the 'phase problem' of crystallography. The situation for the non-crystalline specimen is different, however, in that the pattern is continuous rather than limited to

discrete Bragg peaks. This continuous pattern can therefore be sampled on a finer scale. That sufficient oversampling can lead to a reconstruction was pointed out by Bates⁴. To perform such a reconstruction, Chapman² devised a Fienup-type¹⁷ iterative algorithm. Using a strengthened form of this, Miao *et al.*⁵ were able not only to perform reconstructions of model data in two and three dimensions, but also to show that the degree of oversampling called for by Bates⁴ can be relaxed somewhat for the higher-dimensional cases.

In our experiment we made use of this reconstruction algorithm. The reconstruction from the diffraction pattern of Fig. 2 is shown in Fig. 4. Our phasing algorithm uses knowledge of a finite support which is defined as an enclosing boundary of the specimen. In this reconstruction, we chose a $5.7\ \mu\text{m} \times 5.7\ \mu\text{m}$ square as the finite support which is larger than the size of the image itself. The initial input to the iterative algorithm was a random phase set and, after about 1,000 iterations, a good reconstruction (Fig. 4) was obtained. The computing time of 1,000 iterations is ~ 30 min on a 450-MHz Pentium II workstation. Details of the reconstruction procedure are given elsewhere^{5,16}. The reconstructed image is consistent with the resolution limit, ~ 75 nm, set by the angular extent of the CCD detector. The inner portion of the diffraction pattern could also be filled by Fourier processing of a moderate-resolution image of the specimen made with a scanning transmission X-ray microscope¹, whereupon a reconstruction with an almost perfectly clean background was obtained.

We believe that the successful recording and reconstruction of the test pattern reported here is the critical step that will open the way to high-resolution three-dimensional imaging of such structures as small whole cells, or large sub-cellular structures, in cell biology. Extension from two to three dimensions requires that a series of diffraction patterns be recorded as the specimen is rotated around an axis perpendicular to the beam. We have taken the first steps in this direction. Model calculations indicate that the iterative algorithm used in this work is able to reconstruct such a data set⁵. To be able to collect the data set from a biological (or other radiation-sensitive) specimen, it would be necessary to keep the specimen near the temperature of liquid nitrogen. Experiments show that specimens at this temperature can withstand a radiation dose up to 10^{10} Gy without observable morphological damage^{18,19}. Finally, to improve the resolution without sacrificing specimen size, a CCD detector with more pixels would be needed: such detectors are now commercially available. □

Received 24 March; accepted 8 June 1999.

- Kirz, J., Jacobsen, C. & Howells, M. Soft X-ray microscopes and their biological applications. *Q. Rev. Biophys.* **28**, 33–130 (1995).
- Sayre, D. & Chapman, H. N. X-ray microscopy. *Acta Crystallogr.* **A 51**, 237–252 (1995).
- Millane, R. P. Phase retrieval in crystallography and optics. *J. Opt. Soc. Am.* **A 7**, 394–411 (1990).
- Bates, R. H. T. Fourier phase problems are uniquely solvable in more than one dimension. I: underlying theory. *Optik* **61**, 247–262 (1982).
- Miao, J., Sayre, D. & Chapman, H. N. Phase retrieval from the magnitude of the Fourier transforms of non-periodic objects. *J. Opt. Soc. Am.* **A 15**, 1662–1669 (1998).
- Sayre, D., Kirz, J., Feder, R., Kim, D. M. & Spiller, E. Potential operating region for ultrafast X-ray microscopy of biological specimens. *Science* **196**, 1339–1340 (1977).
- Jacobsen, C. & Kirz, J. X-ray microscopy with synchrotron radiation. *Nature Struct. Biol.* **5**, (synchrotron suppl.), 650–653 (1998).
- Jacobsen, C., Kirz, J. & Williams, S. Resolution in soft X-ray microscopes. *Ultramicroscopy* **47**, 55–79 (1992).
- Thieme, J., Schmahl, G., Umbach, E. & Rudolph, D. (eds) *X-ray Microscopy and Spectromicroscopy* (Springer, Berlin, 1998).
- Haddad, W. S. *et al.* Ultra high resolution x-ray tomography. *Science* **266**, 1213–1215 (1994).
- Lehr, L. 3D x-ray microscopy: tomographic imaging of mineral sheaths of bacteria *Leptothrix ochracea* with the Göttingen x-ray microscope at BESSY. *Optik* **104**, 166–170 (1997).
- Wang, Y., Jacobsen, C., Maser, J. & Osanna, A. Soft x-ray microscopy with cryo STXM: II. Tomography. *J. Microsc.* (in the press).
- Howells, M. *et al.* X-ray holograms at improved resolution: a study of zymogen granules. *Science* **238**, 514–517 (1987).
- Lindaas, S., Howells, M., Jacobsen, C. & Kalinovsky, A. X-ray holographic microscopy by means of photoresist recording and atomic-force microscope readout. *J. Opt. Soc. Am.* **A 13**, 1788–1800 (1996).
- Sayre, D. in *Imaging Processes and Coherence in Physics* (eds Schlenker, M. *et al.*) 229–235 (Springer, Berlin, 1980).
- Sayre, D., Chapman, H. N. & Miao, J. On the extendibility of X-ray crystallography to noncrystals. *Acta Crystallogr.* **A 54**, 233–239 (1998).
- Fienup, J. R. Phase retrieval algorithm: a comparison. *Appl. Opt.* **21**, 2758–2769 (1982).
- Schneider, G. & Niemann, B. in *X-ray Microscopy and Spectromicroscopy* (eds Thieme, J., Schmahl, G., Rudolph, D. & Umbach, E.) 25–34 (Springer, Berlin, 1998).

- Maser, J. *et al.* in *X-ray Microscopy and Spectromicroscopy* (eds Thieme, J., Schmahl, G., Rudolph, D. & Umbach, E.) 35–44 (Springer, Berlin, 1998).
- Lindaas, S. *et al.* in *X-ray Microscopy and Spectromicroscopy* (eds Thieme, J., Schmahl, G., Rudolph, D. & Umbach, E.) 75–86 (Springer, Berlin, 1998).

Acknowledgements. The decision to try oversampling as a phasing technique was arrived at in a conversation in the late 1980s with G. Bricogne. W. Yun and H. N. Chapman also participated in early parts of this experiment. We thank C. Jacobsen for help and advice, especially with the numerical reconstruction, and we thank him and M. Howells for use of the apparatus²⁰ in which the exposures were made; we also thank S. Wirick for help with data acquisition. P.C. thanks the Leverhulme Trust Great Britain for supporting the nanofabrication programme at King's College, London. This work was performed at the National Synchrotron Light Source, which is supported by the US Department of Energy. Our work was supported in part by the US Department of Energy.

Correspondence and requests for materials should be addressed to J.M. (e-mail: miao@xray1.physics.sunysb.edu).

Forcing of the cold event of 8,200 years ago by catastrophic drainage of Laurentide lakes

D. C. Barber*, A. Dyke†, C. Hillaire-Marcel‡, A. E. Jennings*, J. T. Andrews*, M. W. Kerwin*, G. Bilodeau‡, R. McNeely†, J. Southon§, M. D. Morehead* & J.-M. Gagnon||

* Institute for Arctic & Alpine Research, and Department of Geological Sciences, University of Colorado, Boulder, Colorado 80309, USA

† Geological Survey of Canada, 601 Booth Street, Ottawa K1A 0E8, Canada

‡ Centre de recherche en géochimie isotopique et en géochronologie, Université du Québec à Montréal, Québec H3C 3P8, Canada

§ Center for Accelerator Mass Spectrometry, L-397, Lawrence Livermore National Laboratory, PO Box 808, Livermore, California 94551, USA

|| Canadian Museum of Nature, PO Box 3443, Station D, Ottawa, Ontario K1P 6P4, Canada

The sensitivity of oceanic thermohaline circulation to freshwater perturbations is a critical issue for understanding abrupt climate change¹. Abrupt climate fluctuations that occurred during both Holocene and Late Pleistocene times have been linked to changes in ocean circulation^{2–6}, but their causes remain uncertain. One of the largest such events in the Holocene occurred between 8,400 and 8,000 calendar years ago^{2,7,8} (7,650–7,200 ¹⁴C years ago), when the temperature dropped by 4–8 °C in central Greenland² and 1.5–3 °C at marine^{4,7} and terrestrial^{7,8} sites around the north-eastern North Atlantic Ocean. The pattern of cooling implies that heat transfer from the ocean to the atmosphere was reduced in the North Atlantic. Here we argue that this cooling event was forced by a massive outflow of fresh water from the Hudson Strait. This conclusion is based on our estimates of the marine ¹⁴C reservoir for Hudson Bay which, in combination with other regional data, indicate that the glacial lakes Agassiz and Ojibway^{9–11} (originally dammed by a remnant of the Laurentide ice sheet) drained catastrophically $\sim 8,470$ calendar years ago; this would have released $>10^{14}$ m³ of fresh water into the Labrador Sea. This finding supports the hypothesis^{2,7,8} that a sudden increase in freshwater flux from the waning Laurentide ice sheet reduced sea surface salinity and altered ocean circulation, thereby initiating the most abrupt and widespread cold event to have occurred in the past 10,000 years.

During the period of deglaciation that preceded the abrupt climate event of 8,400–8,000 calendar years (cal. yr) ago (the ‘8.2-kyr event’), a remnant Laurentide ice mass occupied Hudson Bay and served as an ice dam for glacial lakes Agassiz and Ojibway^{9–12} (Fig. 1). The rapid collapse of ice in Hudson Bay allowed lakes Agassiz and Ojibway, which had previously discharged over spillways southeastwards to the St Lawrence estuary, to drain swiftly northwards through the Hudson Strait to the Labrador Sea^{10,13–15}.

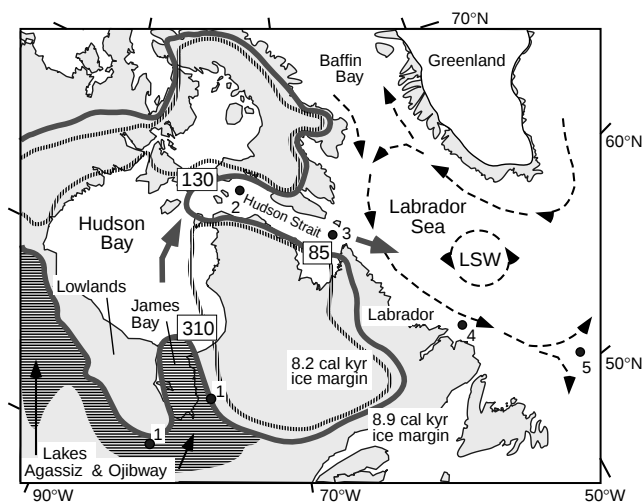


Figure 1 Northeastern Canada and adjacent seas. Former ice-sheet margins¹² are shown for ~8.9 cal. kyr ago (vertical-hatched line) and for ~8.2 cal. kyr ago (thick grey line), before and after disintegration of ice in central Hudson Bay, respectively. Horizontal hatching shows lakes Agassiz and Ojibway^{9–12}. Northward drainage through Hudson Bay and Hudson Strait (dark grey arrows) occurred as the Hudson Bay ice mass disintegrated. Arrows with dashed lines show Labrador Sea current patterns and the area of Labrador Sea Intermediate Water (LSW) formation. Numbers in boxes are regional mean ΔR values (years), based on radiocarbon analyses of mollusc shells collected alive before 1955 ($\Delta R_{\text{local}} = {}^{14}\text{C age}_{\text{meas.}} - \text{contemporaneous surface ocean } {}^{14}\text{C age}$; Table 2). Sites discussed in text and referred to in Table 1 (filled circles) are as follows: site 1, SW and east of James Bay, marine deposits post-dating drainage of glacial lakes Agassiz and Ojibway^{9–16}; site 2, west Hudson Strait cores HU85027-068²¹, 90023-085¹⁴, -099¹⁴, -101¹⁵; site 3, east Hudson Strait cores HU85027-057²¹, 90023-045¹⁵, -064¹⁴, 93034-004¹⁷; site 4, Cartwright saddle¹⁹ cores HU87033-017, -018; and site 5, Orphan knoll²⁰ cores HU91045-094 and MD95-2024.

Before its demise, the Hudson Bay ice mass and the associated proglacial lakes contained a combined volume¹¹ estimated at $5 \times 10^{14} \text{ m}^3$; however, >50% of this was ice that could not have left the Hudson Strait as rapidly as water from the lakes. So to calculate the peak freshwater flux to the Labrador Sea, we consider only the water in lakes Agassiz and Ojibway. The approximate volume of Lake Ojibway^{9,11} before its abrupt northward drainage was 10^{14} m^3 . The volume of Lake Agassiz at that time is not well constrained, but we follow Veilleux¹¹ in assuming that the volume approximately equalled that of Ojibway. Thus the total freshwater volume released by drainage of both lakes is $\sim 2 \times 10^{14} \text{ m}^3$. Before drainage, the lake surfaces stood $\geq 175 \text{ m}$ above the contemporaneous sea level^{9,11}, providing a large initial hydraulic head that drove the outburst once a conduit to the sea opened.

The stratigraphy of the Hudson and James Bay lowlands^{9,11,16} (Fig. 1) ubiquitously records the abrupt lake drainage: glacial-marine sediments lie directly above the proglacial lake sediments. Additional evidence for the lake outburst is the 5–80-cm-thick, red-coloured, haematite-rich sediment layer traceable for 700 km in cores from the western to the eastern reaches of the Hudson Strait^{14,15} (Fig. 1). Regional stratigraphic correlations and provenance studies¹⁵ suggest that the Hudson Strait ‘red bed’ shared a common source with red glacial deposits in north-central Hudson Bay and red-brown glaciolacustrine sediments in the former Agassiz and Ojibway basins^{9–11,16}. The simultaneous deposition of the red bed throughout the Hudson Strait at a time in the glacial period when the strait was free of ice^{14,17} required an extraordinary sediment transport mechanism; the catastrophic outburst flood released from lakes Agassiz and Ojibway probably provided this mechanism.

We evaluated the outburst drainage model, which would have resulted in high freshwater flux and long-range sediment dispersal, by running numerical simulations of plume deposition in the Hudson Strait. The discharge resulting from instantaneous removal of the ice dam was estimated, and we then used an oceanic plume sedimentation model¹⁸ to predict the distribution of sediment

Table 1 Dates on drainage of lakes Agassiz and Ojibway

Sample site	Interval (cm)	Laboratory no.	Ref.*	Lat., lon.† (°N, °W)	Radiocarbon age‡ (${}^{14}\text{C yr BP} \pm 1\sigma$)	ΔR § (yr)	Cal. age (cal. yr BP)	1 σ range¶ (cal. yr BP)
SE Hudson Bay								
Post-dates drainage; dates marine incursion								
SE James Bay		Qu 122	9	53.35, 77.57	8,280 $\pm$ 160			
SE James Bay		Qu 124	9	53.35, 77.57	8,150 $\pm$ 180			
SW James Bay		GSC 897	16	50.22, 84.30	8,160 $\pm$ 160#			
SW James Bay		GSC 880	16	51.93, 84.53	8,120 $\pm$ 140#			
					(8,150 $\pm$ 50) ²²	310	8,280	8,330–8,160
West Hudson Strait								
Post-dates drainage; above red bed								
90023-085	98–100	TO 3265	14	62.62, 76.38	8,170 $\pm$ 140	130	8,420	8,600–8,320
Pre-dates drainage; below red bed								
90023-101	365–367	AA 12888	15	63.05, 74.30	8,260 $\pm$ 60			
90023-099	320–325	AA 12887	14	63.07, 74.57	8,270 $\pm$ 70			
85027-068	989–996	TO 751	21	63.08, 74.31	8,310 $\pm$ 70			
					(8,280 $\pm$ 40)	130	8,550	8,650–8,490
East Hudson Strait								
Post-dates drainage; above red bed								
93034-004	19–21	CAMS 25762	17	61.22, 66.43	8,030 $\pm$ 60			
90023-045	480–483	AA 17380	15	60.95, 66.14	8,155 $\pm$ 130			
90023-064	460–462	TO 3263	14	61.13, 70.58	8,160 $\pm$ 150			
					(8,065 $\pm$ 50)	85	8,380	8,440–8,325
Pre-dates drainage; below red bed								
85027-057	814–822	TO 749	21	61.07, 66.43	8,140 $\pm$ 70			
93034-004	78–80	AA 13055	17	61.22, 66.43	8,395 $\pm$ 70			
90023-045	777–779	AA 11879	15	60.95, 66.14	8,490 $\pm$ 200			
					(9,280 $\pm$ 50)	85	8,610	8,740–8,520

* References cited provide additional sample information.

† Positions in decimal degrees; also see Fig. 1.

‡ Radiocarbon dates given as conventional, $\delta^{13}\text{C}$ -normalized ${}^{14}\text{C}$ ages (no reservoir correction).

§ See Table 2, Fig. 1 and text for derivation of local ΔR values.

|| Calibrated date (cal. yr) converted²³ from weighted mean ${}^{14}\text{C}$ age using local ΔR values. Mean calibrated age between bounding dates on lake drainage is 8,470 cal. yr BP.

¶ Note that calibrated age ranges are asymmetrical due to the nonlinear ${}^{14}\text{C}$ calibration curve.

In accord with laboratory protocol, GSC dates are reported here with 2σ errors; corresponding 1σ errors were used when calculating weighted means from these dates.

* Parentheses enclose weighted averages of preceding ${}^{14}\text{C}$ dates.

Table 2 ^{14}C ages of live-collected Hudson Bay shells

Laboratory no.	Collection site (lat. °N, lon. °W)*	Collection year (AD)	Radiocarbon age (years $\pm 1\sigma$)†	Model age (years)‡	ΔR (years)§
East Hudson Strait and Ungava Bay					
CAMS-33148	60.41, 64.83	1948	630 $\pm$ 50	480	150
CAMS-34644	59.22, 65.75	1947	540 $\pm$ 50	480	60
CAMS-34654	59.48, 65.25	1950	650 $\pm$ 50	480	170
CAMS-46546	61.63, 71.97	1920	500 $\pm$ 50	460	40
CAMS-46550	60.83, 69.93	1950	620 $\pm$ 40	480	140
CAMS-46555	60.07, 69.43	1949	430 $\pm$ 40	480	-50
GSC-6107	59.22, 65.75	1947	480 $\pm$ 40	480	0
TO-5980	59.22, 65.75	1947	650 $\pm$ 40	480	170
$n = 8$		Mean:	560		85
West Hudson Strait and North Hudson Bay					
CAMS-33144	64.40, 77.93	1953	760 $\pm$ 50	480	280
CAMS-33146	66.47, 86.20	1955	690 $\pm$ 50	480	210
CAMS-33149	63.00, 82.65	1954	690 $\pm$ 50	480	210
CAMS-34647	63.60, 82.00	1953	480 $\pm$ 50	480	0
CAMS-34648	64.33, 75.58	1954	600 $\pm$ 50	480	120
CAMS-46547	62.95, 81.84	1954	510 $\pm$ 40	480	30
CAMS-46549	63.00, 82.65	1954	590 $\pm$ 40	480	110
CAMS-46551	64.23, 76.55	1954	430 $\pm$ 40	480	-50
CAMS-46552	63.68, 80.20	1953	560 $\pm$ 40	480	80
CAMS-46556	64.23, 76.55	1954	590 $\pm$ 40	480	110
CAMS-46557	63.00, 82.65	1954	530 $\pm$ 50	480	50
CAMS-46559	64.23, 76.55	1954	520 $\pm$ 50	480	40
CAMS-46560	63.60, 82.00	1953	670 $\pm$ 50	480	190
CAMS-47241	66.92, 81.33	1955	810 $\pm$ 40	480	330
CAMS-47244	62.98, 82.69	1954	610 $\pm$ 40	480	130
TO-5977	64.40, 77.93	1953	690 $\pm$ 50	480	210
$n = 16$		Mean:	610		130
SE Hudson Bay and James Bay					
CAMS-46545	56.50, 77.00	1920	630 $\pm$ 40	460	170
CAMS-46561	56.25, 76.33	1920	580 $\pm$ 50	460	120
CAMS-46755	52.00, 79.50	1941	970 $\pm$ 40	475	495
CAMS-46757	52.95, 79.00	1920	720 $\pm$ 40	460	260
CAMS-46759	52.00, 79.50	1920	1,050 $\pm$ 40	460	590
CAMS-46760	52.95, 79.00	1920	740 $\pm$ 40	460	280
CAMS-46761	52.60, 78.75	1920	810 $\pm$ 40	460	350
CAMS-47247	55.28, 77.75	1949	560 $\pm$ 50	480	80
CAMS-48978	53.12, 79.86	1920	740 $\pm$ 40	460	280
CAMS-48979	52.00, 79.50	1920	940 $\pm$ 40	460	480
$n = 10$		Mean:	775		210

* Live shell sample locations in decimal degrees.

† Laboratory-reported, conventional, $\delta^{13}\text{C}$ -normalized ^{14}C ages.

‡ Model-derived mean surface ocean ages for the year of collection²².

§ ΔR = measured shell radiocarbon age minus modelled surface ocean age. Regional means of the ΔR values are used in calibration²⁶ of ages for the lake outburst event (Table 1).

resulting from an event of this magnitude. The simulation predicted deposit thicknesses of ~ 50 cm in the western Hudson Strait, thinning to ~ 30 cm in the eastern Hudson Strait. The approximate agreement of modelled deposit thicknesses with the observed red bed^{14,15,17} supports the interpretation that final drainage of lakes Agassiz and Ojibway produced a massive freshwater pulse to the Labrador Sea (Fig. 1).

Evidence for a freshwater pulse is also found beyond the Hudson Strait. Cores from Cartwright saddle on the Labrador shelf (Fig. 1) exhibit a 0.6‰ reduction in the $\delta^{18}\text{O}$ of planktonic foraminifera between 8.5 and 8.3 cal. kyr ago ($\sim 7.8\text{--}7.6$ ^{14}C kyr)¹⁹. Farther southeast, piston cores from the flank of Orphan knoll (Fig. 1) show increased offsets between the $\delta^{18}\text{O}$ compositions of planktonic foraminifera *Globigerina bulloides* and left-coiling *Neogloboquadrina pachyderma* (shallow- and deep-dwelling species, respectively)²⁰. The offset increases from 0.8‰ before the freshwater pulse to 1.3‰ during the pulse, due to a 0.5‰ reduction in the *G. bulloides* values. This isotopic shift is comparable to that at Cartwright saddle and indicates lower sea surface salinities and increased water-mass stratification throughout the western Labrador Sea²⁰. Thus data from various sites support a large freshwater plume entering the Labrador Sea in the early Holocene. We determined the calendar age of the Laurentide lake outburst to ascertain whether this freshwater input coincided with onset of the '8.2-kyr' cold event.

The published radiocarbon dates constraining the final Agassiz–Ojibway drainage are on marine carbonates from the Hudson and James Bay lowlands^{9,12,16} and the Hudson Strait^{14,15,17,21} (Table 1; Fig. 1). Because of the pattern of ice marginal recession along the

southern margin of Hudson Bay, the post-glacial sea reached its highstand throughout that region simultaneously. Dates on fossil marine bivalves from uplifted basal post-glacial marine sediments in the Hudson Bay lowlands constrain the time of initial marine incursion, and thus also provide limiting ages that post-date drainage of lakes Agassiz and Ojibway (Table 1; Fig. 1)^{9,16}. In the Hudson Strait (Fig. 1), the radiocarbon ages bounding the freshwater pulse are on bivalves and foraminifera collected above and below the red bed in marine sediment cores (Table 1; Fig. 1). In earlier work on the overall chronostratigraphy of the Hudson Strait^{14,15,17,21}, many more dates were obtained than we include in Table 1. We excluded many dates because sediment reworking obfuscated their stratigraphic context. Of the remaining dates, we consider only those that most closely constrain deposition of the red bed.

Conversion of ^{14}C dates to the calendar-year timescale allows comparison with events in ice-core chronologies^{2,6} produced by counting of annual layers, but precise calibration cannot be performed without a correction for the local marine ^{14}C reservoir effect²². The local reservoir correction is the sum, in radiocarbon years, of the mean global surface ocean reservoir age (R) plus any local deviation (ΔR) from the contemporaneous global-mean ocean age²². Detailed comparisons between radiocarbon dates from different regions or reservoirs (for example, the atmosphere) require an estimate of ΔR . In the past, ΔR values were not well established; so in earlier work, reservoir corrections of 400–450 years (that is, $\Delta R = 0$) were applied to shallow marine dates. Here we estimate local ΔR values using radiocarbon dates on 34 museum shell

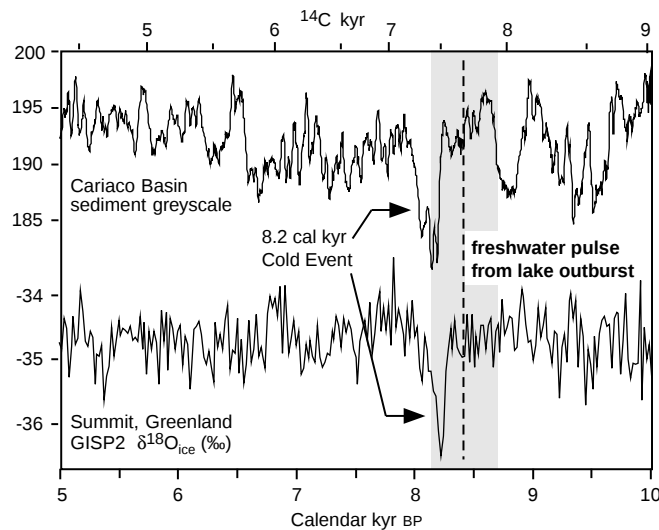


Figure 2 Climate proxy records of the '8.2-kyr' cold event. Both ^{14}C (top) and calendar (lower) timescales are given. Upper curve shows Cariaco basin greyscale record; reduced greyscale values indicate increased zonal wind speed due to high-latitude cooling⁵. Timescale of greyscale variations differs from that in ref. 5 due to subsequent work by those authors; data on the revised timescale are available from the World Data Center-A for Paleoclimatology (<http://www.ngdc.noaa.gov/paleo>).

Lower curve shows bidecadal $\delta^{18}\text{O}$ values of ice from the GISP2 ice core⁶, interpreted to reflect primarily the temperature of precipitation over Summit, Greenland; more negative values indicate colder temperatures. Also shown is age for the lake drainage event: 8,470 cal. yr BP or ~ 7.7 ^{14}C kyr (vertical dashed line); extremes of the 1σ cal. age ranges on the bounding dates (Table 1) give an error range of 8,160–8,740 cal. yr BP (shaded).

specimens that were collected alive in the Hudson Bay region from AD 1920 to 1955, before significant contamination of the atmosphere with bomb radiocarbon (Table 2). The ages of these shells range from 430 to 1,050 ^{14}C yr (1σ errors range from 40 to 50 yr). Despite the observed variability, the ages typically exceed 550 yr and the means of ages from each region increase consistently with distance from the open Labrador Sea (Fig. 1). By comparing the shell dates with modelled mean-ocean reservoir age data²², we derive ΔR values for the southeastern Hudson Bay and James Bay of 310 ± 50 yr, for the northern Hudson Bay and western Hudson Strait of 130 ± 50 yr, and for the eastern Hudson Strait of 85 ± 50 yr (Table 2; Fig. 1).

We identify two possible causes of the higher ΔR values (that is, lower initial ^{14}C activities) for the marine carbon pools in the Hudson Bay and Hudson Strait: (1) runoff draining Palaeozoic limestone bedrock and carbonate-rich glacial sediments in the region provides dissolved ^{14}C -free bicarbonate²³; (2) persistent sea-ice coverage inhibits air–sea ^{14}C equilibration. Modelling²⁴ suggests that the 7–8 months of seasonal sea ice in the Hudson Strait²⁵ could cause an apparent ageing of the marine ^{14}C reservoir by 150–200 yr, slightly more than observed (Table 2; Fig. 1). However, despite a shorter sea-ice season, James Bay shells have the highest ΔR values in the region (Table 2; Fig. 1). Significant runoff from carbonate-rich areas enters James Bay, thus both sea ice and inputs of ancient carbon seem to contribute to the observed pattern in ΔR values. To calibrate ^{14}C dates from the deglacial period 8.8–7.5 cal. kyr ago, we must evaluate possible differences in the factors influencing ΔR . During the time of interest, high inputs of melt water from the residual ice sheet¹² (Fig. 1) probably lengthened the sea-ice season²⁵. Additionally, ^{14}C -free bicarbonate input was probably higher due to the abundance of fresh, glacially abraded Palaeozoic carbonate²³. The effects on ΔR of differing water masses and current patterns during deglaciation are not known, although the lower percentage of Labrador Sea surface water (pre-bomb $\Delta R = 0$) in the mixed Hudson Bay water mass²⁵ may have produced larger ΔR values during deglaciation. Taken together, these effects imply that the regional ΔR values in Table 2 are conservative (that is, low) with respect to those $\sim 8,000$ years ago.

The ΔR values derived here facilitate conversion of radiocarbon ages for the final Agassiz–Ojibway drainage into calendar ages using

the marine calibration scheme in CALIB 3.03A²⁶. We calculated the age of the freshwater pulse (8,470 cal. yr before present, BP) as the midpoint between means of both the younger and older event-bounding calibrated ages (Table 1). Within the limits of annual ice-core layer counting, radiocarbon dating, ΔR estimates, and ^{14}C -to-calendar year conversion, the age derived here for final northward drainage of lakes Agassiz and Ojibway coincides with the $8,400 \pm 100$ cal. yr BP onset of climate cooling in Greenland and elsewhere^{2–8} (Fig. 2).

The cataclysmic release of $2 \times 10^{14} \text{ m}^3$ of lake water over 1, 10 or 100 years would have increased the freshwater flux to the Labrador Sea by 6, 0.6 or 0.06 Sv ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$), respectively. Numerical simulation of the Hudson Strait redbed deposit suggests that drainage occurred in less than one year, but the available chronology does not yield a precise duration. Results from ocean circulation models^{27–29} suggest that excess freshwater discharges of 0.06–0.12 Sv can reduce the formation rates of Labrador Sea Intermediate Water (LSW) and North Atlantic Deep Water (NADW), thereby strongly affecting ocean heat transport. These simulations do not specifically apply to the lake outburst case, however, because the excess discharges were prescribed for periods of >500 years in the models^{27–29}. Although the ocean freshening due to Agassiz–Ojibway drainage was of shorter duration, the lakewater pulse was preceded by an interval (600–900 years) of somewhat reduced sea surface salinity in the Labrador Sea. This previous low-salinity interval, recorded by reduced $\delta^{18}\text{O}$ values and increased ice-rafted detritus at both Cartwright saddle¹⁹ and Orphan knoll²⁰ (Fig. 1), apparently resulted from the advance and subsequent breakup of a partly marine-based northern Labrador ice sheet¹⁷.

The low sea surface salinities resulting from the Agassiz–Ojibway outburst propagated southeast from Hudson Strait, producing more pronounced freshening in the region of LSW formation (Fig. 1) than at the more distant NADW formation sites⁷. The present northward ocean heat transport associated with formation of LSW is 0.3 PW ($1 \text{ PW} = 10^{15} \text{ W}$), half that due to formation of NADW (0.6 PW)³⁰. If the formation of both LSW and NADW ceased during the Younger Dryas cold event, but only LSW formation was disrupted during the '8.2-kyr' event, then for the latter event we might expect regional atmospheric cooling of one-third the magnitude as that during the Younger Dryas. This scenario

undoubtedly oversimplifies ocean circulation and climate boundary conditions, but the resulting prediction of the relative amplitudes of the two cold events approximates the relative cooling observed in proxy records that contain both events^{2–6}.

Evidence presented here—of a large freshwater pulse from the final outburst drainage of lakes Agassiz and Ojibway, together with the revised timing of this pulse ($\sim 8,470$ cal. yr BP, or ~ 7.7 ^{14}C kyr BP)—directly supports the hypothesis^{3,7,8} that an increase in freshwater flux modified ocean circulation, thereby causing the observed '8.2-kyr' climate cooling. This result provides perspective both on the sensitivity of ocean circulation to freshwater inputs and on the climate oscillations of the present interglacial period. Specifically, our findings suggest that in the case of the '8.2-kyr' event, the thermohaline circulation responded to exceptionally strong forcing: initiation of the most abrupt and widespread climate shift known from the past 10,000 (calendar) years required a massive, albeit short-lived, perturbation of the North Atlantic freshwater balance. □

Received 26 March; accepted 8 June 1999.

1. Broecker, W. S. Thermohaline circulation, the Achilles heel of our climate system: Will man-made CO_2 upset the current balance? *Science* **278**, 1582–1588 (1997).
2. Alley, R. B. *et al.* Holocene climatic instability: a prominent, widespread event 8200 yr ago. *Geology* **25**, 483–486 (1997).
3. Street-Perrott, F. A. & Perrott, R. A. Abrupt climate fluctuations in the tropics: the influence of the Atlantic Ocean circulation. *Nature* **343**, 607–612 (1990).
4. Bond, G. *et al.* A pervasive millennial-scale cycle in North Atlantic Holocene and Glacial climates. *Science* **278**, 1257–1266 (1997).
5. Hughen, K. A., Overpeck, J. T., Trumbore, S. & Peterson, L. C. Rapid climate changes in the tropical Atlantic region during the last deglaciation. *Nature* **380**, 51–54 (1996).
6. Stuiver, M., Grootes, P. M. & Braziunas, T. F. The GISP2 ^{18}O record of the past 16,500 yrs and the role of the sun, ocean and volcanoes. *Quat. Res.* **44**, 341–354 (1995).
7. Klitgaard-Kristensen, D., Sejrup, H. -P., Hafliðason, H., Johnsen, S. & Spurk, M. A regional 8200 cal. yr BP cooling event in northwest Europe, induced by final stages of the Laurentide ice-sheet deglaciation? *J. Quat. Sci.* **13**, 165–169 (1998).
8. von Grafenstein, U., Erlenkeuser, H., Müller, J., Jouzel, J. & Johnsen, S. The cold event 8200 years ago documented in oxygen isotope records of precipitation in Europe and Greenland. *Clim. Dyn.* **14**, 73–81 (1998).
9. Hardy, L. La déglaciation et les épisodes lacustre et marin sur les versants de la partie québécoise des basses terres de la baie de James. *Geogr. Phys. Quat.* **31**, 261–273 (1977).
10. Teller, J. T. In *The Geology of North America* Vol. K-3, *North America and Adjacent Oceans during the Last Deglaciation* (eds Ruddiman, W. F. & Wright, H. E. Jr) 39–69 (Geol. Soc. of America, Boulder, Colorado, 1987).
11. Veillette, J. J. Evolution and paleohydrology of Glacial Lakes Barlow and Ojibway. *Quat. Sci. Rev.* **13**, 945–971 (1994).
12. Dyke, A. S. & Prest, V. K. Paleogeography of northern North America, 18000–5000 years ago. (Map 1703A, Scale 1:12500000, Geol. Surv. of Canada, Ottawa, 1989).
13. Andrews, J. T. & Falconer, G. Late glacial and postglacial history and emergence of the Ottawa Islands, Hudson Bay, N.W.T.: Evidence on the deglaciation of Hudson Bay. *Can. J. Earth Sci.* **6**, 1263–1276 (1969).
14. Andrews, J. T. *et al.* Final stages in the collapse of the Laurentide Ice Sheet, Hudson Strait, Canada, NWT: Based on ^{14}C AMS dates and magnetic susceptibility logs. *Quat. Sci. Rev.* **14**, 983–1004 (1995).
15. Kerwin, M. W. A regional stratigraphic isochron (ca. 8000 ^{14}C yr B.P.) from the final deglaciation of Hudson Strait. *Quat. Res.* **46**, 89–98 (1996).
16. Skinner, R. G. Quaternary stratigraphy of the Moose River basin, Ontario. *Geol. Surv. Can. Bull.* **225** (1973).
17. Jennings, A. E., Manley, W. F., MacLean, B. & Andrews, J. T. Marine evidence for the last glacial advance across eastern Hudson Strait, eastern Canadian Arctic. *J. Quat. Sci.* **13**, 501–514 (1998).
18. Syvitski, J. P., Skene, K. I., Nicholson, M. K. & Morehead, M. D. Plume 1. 1: Deposition of sediment from a fluvial plume. *Comput. Geosci.* **24**, 159–171 (1998).
19. Andrews, J. T., Keigwin, L., Hall, F. R. & Jennings, A. E. Late Quaternary (12 ka) sediment and meltwater events on the Labrador shelf: Evidence from high-resolution cores in Cartwright Saddle (54–55°N). *J. Quat. Sci.* (in the press).
20. Bilodeau, G., Hillaire-Marcel, C., de Vernal, A. & Stoner, J. Changes in vertical structure of Labrador Sea water masses during the last 25 ka based on oxygen isotopes in planktic and benthic foraminifera. *Geosci. Can.* **23**, (Suppl. Quebec 1998 Abstr.) 18–19 (1998).
21. Vilks, G., MacLean, B., Deonarine, B., Currie, C. G. & Moran, K. Late Quaternary paleoceanography and sedimentary environments in Hudson Strait. *Geogr. Phys. Quat.* **43**, 161–178 (1989).
22. Stuiver, M. & Braziunas, T. F. Modeling atmospheric ^{14}C influences and ^{14}C ages of marine samples to 10,000 BC. *Radiocarbon* **35**, 137–189 (1993).
23. Karrow, P. F. Carbonates, granulometry, and color of tills on the south-central Canadian Shield and their implications for stratigraphy and radiocarbon dating. *Boreas* **21**, 379–391 (1992).
24. Bard, E. *et al.* The North Atlantic atmosphere-sea surface ^{14}C gradient during the Younger Dryas climatic event. *Earth Planet. Sci. Lett.* **126**, 275–287 (1994).
25. Drinkwater, K. F. In *Canadian Inland Seas* (ed. Martini, I. P.) 237–264 (Elsevier, New York, 1986).
26. Stuiver, M. & Reimer, P. J. Radiocarbon calibration program Rev 3.0.3A - Mac Test Version #6. *Radiocarbon* **35**, 215–230 (1993).
27. Rahmstorf, S. On the freshwater forcing and transport of the Atlantic thermohaline circulation. *Clim. Dyn.* **12**, 799–811 (1996).
28. Fanning, A. F. & Weaver, A. J. Temporal-geographical meltwater influences on the North Atlantic Conveyor: Implication for the Younger Dryas. *Paleoceanography* **12**, 307–320 (1997).
29. Manabe, S. & Stouffer, R. J. Coupled ocean-atmosphere model response to freshwater input: Comparison to Younger Dryas event. *Paleoceanography* **12**, 321–336 (1997).
30. Talley, L. D. In *Mechanisms of Millennial-scale Global Climate Change*. (eds Clark, P. U., Webb, R. & Keigwin, L.) (AGU Monogr., Am Geophys. Union, Washington DC, in the press).

Acknowledgements. We thank the Canadian Museum of Nature for providing archived live-collected shells, and G. Bond, D. Fisher, B. MacLean and J. Teller for comments on the manuscript. This work was supported by the Terrain Sciences Division, Geological Survey of Canada, and the US NSF (A.E.J. and J.T.A.).

Correspondence and requests for materials should be addressed to D.B. (e-mail: barberdc@ucsub.colorado.edu)

Asynchronous deposition of ice-rafted layers in the Nordic seas and North Atlantic Ocean

J. A. Dowdeswell*, A. Elverhøi†, J. T. Andrews‡ & D. Hebbeln§

* Bristol Glaciology Centre, School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK

† Department of Geology, University of Oslo, Postboks 1047, Blindern, N-0316 Oslo, Norway

‡ Institute of Arctic and Alpine Research and Department of Geological Sciences, University of Colorado, Boulder, Colorado 80309, USA

§ FB Geowissenschaften, University of Bremen, Postfach 330440, D-28344 Bremen, Germany

Instabilities in ice-stream flow within the North American Laurentide Ice Sheet, leading to the periodic release of armadas of icebergs into the North Atlantic Ocean over the past 60,000 years, have produced extensive layers of coarse-grained iceberg-rafted debris (Heinrich layers) in North Atlantic sediments^{1,2}. Correlation of these layers with iceberg-discharge events from the ice sheets on Greenland, Iceland and Scandinavia, suggested in previous studies for some Heinrich layers and in some areas^{3–5}, would imply that ice-sheet instability had been synchronous across the North Atlantic, presumably in response to a common environmental cause. Here we show a lack of widespread systematic correlations, both between ice-rafted debris layers in different sediment cores from the Nordic seas, and between the Nordic layers and the North Atlantic Heinrich layers. This suggests that the full-glacial Nordic ice sheets did not exhibit unstable behaviour coincident with iceberg discharge from the vast Hudson Bay drainage basin of the Laurentide Ice Sheet^{6,7}. Off the Hudson Strait, significant ice-sheet discharge of melt water is indicated by size-sorted sandy and muddy turbidite sediments, different from the poorly sorted debris flows which dominate sedimentation on the margins of the Nordic seas^{8–10}. Together, these results suggest that the dynamics of Quaternary ice sheets surrounding the Nordic seas were different from the outlet glacier draining the Hudson Bay basin, and they provide evidence against a common circum-North-Atlantic mechanism driving the discharge of icebergs.

A number of marine geological studies have demonstrated that huge numbers of icebergs, derived mainly from the Hudson Strait, produced a series of six rapidly deposited¹¹ Heinrich layers of iceberg-rafted debris (IRD). These distinctive layers range from about one metre to a few centimetres in thickness, and can be traced for more than 3,000 km across the North Atlantic^{1,2}. However, although many sediment cores have been examined from the Nordic seas (Fig. 1), coarse-grained layers of the thickness and spatial continuity of the Heinrich layers have not been identified. Correlative IRD events have so far been found only in some areas and for some of the Heinrich events^{3–5,12,13} (Fig. 2).

Cores from sites adjacent to outlet glaciers of the Fennoscandian, Svalbard-Barents Sea and Greenland ice sheets (Fig. 1b) show that there is little evidence for a systematic regional correlation of IRD layers across the Nordic seas, nor is there a consistent linkage with the North Atlantic Heinrich layers (Fig. 2). A detailed study of IRD

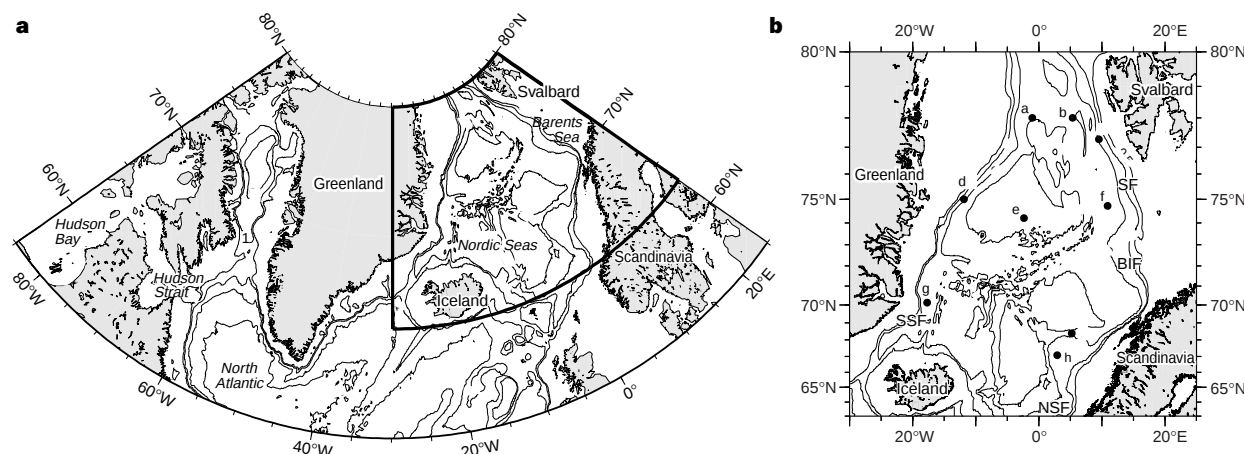


Figure 1 The Nordic seas and the North Atlantic Ocean. **a**, Map of the North Atlantic and Nordic seas with bathymetric contours in thousands of metres. Hudson Bay and the Hudson Strait in Canada are the main source regions for Heinrich layers 1–6¹². **b**, The locations of nine sediment cores in the Nordic seas,

from which records of ice-rafted debris have been derived (labelled a–i). The chief glacier-influenced fans are also shown: SF, Storfjorden fan (35,000 km²); BIF, Bear Island fan (215,000 km²); NSF, the North Sea fan (142,000 km²); and SSF, the Scoresby Sund fan (19,000 km²)^{8,10,16}.

variations in deep-sea cores off the mid-Norwegian shelf identified ten well-defined IRD peaks in the interval from 54,300 to 13,000 calendar years ago¹². Only two of these peaks correlate with Heinrich layers (H2 and H3), and even this assumes an age uncertainty of $\pm 2,000$ years (ref. 12). Thus, detailed correlations may, in some cases, be 'event matching' rather than based on a suite of bracketing AMS (accelerator mass spectrometer) radiocarbon dates¹⁴. In addition, several of the IRD events identified in isotope stages 3 and 4 by Baumann *et al.*¹² do not correspond to the IRD peaks identified by Fronval *et al.*³, although the cores were obtained in nearby areas of the Nordic seas. IRD events in the Nordic seas^{12,15} therefore appear not to represent synchronous responses to pan-North-Atlantic environmental change (Fig. 1a).

The clearest evidence for synchronicity within the IRD records is associated with the effect of major regional deglaciations of Svalbard and Scandinavia, rather than linkage with periodic instability of the Hudson Strait–Hudson Bay ice-sheet drainage basin. These deglaciations resulted in enhanced IRD fluxes and relatively synchronous IRD deposition in the eastern Nordic seas during the periods 55–50,000 years ago and 15–10,000 years ago¹² (Fig. 2).

The general lack of synchronicity in IRD records, both within the Nordic seas and with the North Atlantic Heinrich events (Fig. 2), suggests significant differences in ice dynamics and the nature of sediment delivery between the Hudson Strait outlet of the Laurentide Ice Sheet and those ice-sheet outlet glaciers draining into the Nordic seas. Marked contrasts in the sedimentary processes and environment on the adjacent continental margins provide evidence of long-term differences in the dynamics of these ice-sheet outlets. West of the Eurasian Ice Sheet^{8–10}, and off east Greenland¹⁶, a number of large submarine fans are present on the continental margin (Fig. 1). Each of the fans formed at the mouth of a cross-shelf trough or a major fjord system. The fans have a similar internal architecture and are built up mainly of series of stacked debris flows of heterogeneous grain size, inferred to have been produced as a result of high rates of sediment delivery of subglacial deformation till¹⁷ to the margin from fast-flowing Quaternary ice streams^{8–10}. Similar stacked debris-flow deposits have also been found on the northern Newfoundland, Labrador and Baffin margins of North America, often beyond the edge of cross-shelf troughs draining ice from the full-glacial Laurentide Ice Sheet^{18–20}. As glaciers advanced across Arctic continental shelves during glacial periods, ice was thickest in cross-shelf troughs. Basal melting was most likely beneath thicker ice, and fast-flowing ice streams are inferred to have developed; between ice streams, ice may still have reached the shelf break, but flowed one to two orders of magnitude more slowly

and sedimentation was less rapid^{21,22}. However, the poorly sorted grain size of the debris flows making up the bulk of fan sediments suggests that deposition was not associated with significant melt-water discharge events.

By contrast, sedimentation off the Hudson Strait, which drained the huge Hudson Bay basin of the Laurentide Ice Sheet (between 1.5 and 2.5 million km²)²³, indicates significant ice-sheet discharge of melt water through the presence of size-sorted sandy and muddy turbidite sediments, in addition to the presence of debris-flow deposits^{20,24}. The Labrador Sea slope not only has areas of debris-flow deposition, but also has a well-developed channel system (fed by turbid currents) which delivers predominantly fine-grained muds, derived from glacial meltwater transport and deposition, to the deep-sea northwest Atlantic mid-ocean channel^{20,24,25}.

This large contribution from meltwater-sorted sediments differs from the poorly sorted debris which dominates on the Nordic seas margin^{8–10}. It suggests that the Hudson Strait–Hudson Bay basin of the Laurentide Ice Sheet behaved in a different way from the Quaternary ice sheets draining into the Nordic seas, and even from some of the outlet glaciers draining into other areas of the Canadian margin. Numerical models of the Quaternary ice stream in the Hudson Strait have predicted the initiation of rapid flow through the basal melting of progressively thickening ice on time-scales of thousands of years^{6,7}. This fast flow is sustained for up to several hundred years, with the production of large numbers of icebergs at the mouth of the Hudson Strait, before freezing at the base of the thinning ice sheet curtails rapid motion and the ice stagnates. It is the strongly oscillatory behaviour, and the magnitude of periodic collapse, of the Hudson Strait outlet glacier of the Laurentide Ice Sheet^{6,7} that produces the relatively thick and extensive Heinrich layers.

By contrast, the IRD signal in the Nordic seas, and elsewhere, is likely to have a significant random component, in part a function of relatively local variability in the drift tracks and sediment-release histories of groups of icebergs²⁶. The correlation, or lack of it, of the IRD signal between cores will tend to vary with distance from the source ice-sheet outlet glacier or ice stream and with the magnitude of the calving event. In addition, ice-sheet basins of varying size are likely to respond asynchronously to any single external forcing simply through their differing bathymetry and the response times of drainage basins of varying area and surface-elevation distribution²⁷. Numerical modelling predicts that the main full-glacial ice-sheet drainage basins surrounding the Nordic seas probably ranged between about 10^4 and 10^6 km² in area, with maximum ice thicknesses between 1 and 3 km (ref. 21), yielding response times of

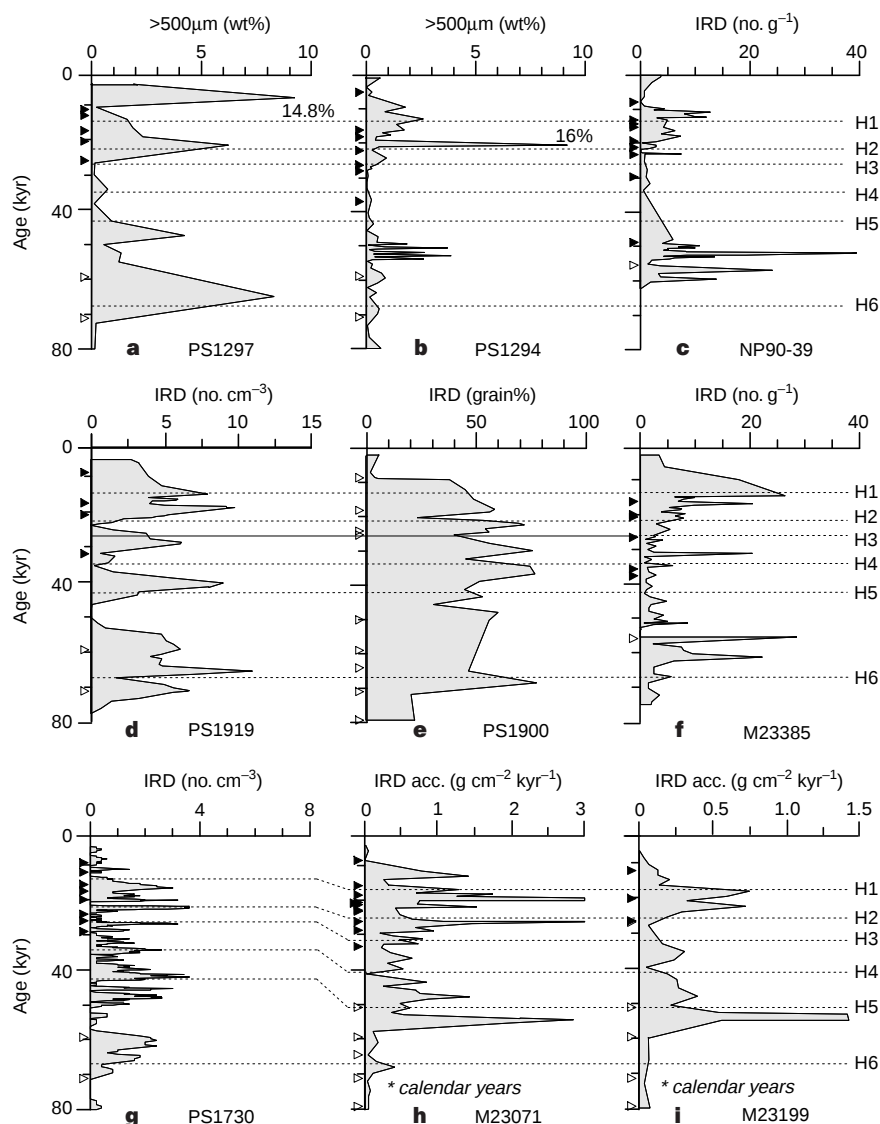


Figure 2 Ice-rafted debris (IRD) records from the Nordic seas. These records indicate asynchronous IRD deposition from the surrounding ice sheets, and little correlation with the North Atlantic Heinrich layers (H1–H6). The cores, located as a–i in Fig. 1b, are plotted on a radiocarbon timescale, except for cores M23071 and M23199, which are in calendar years. Age control by AMS dating (filled triangles) and by $\delta^{18}\text{O}$ data (open triangles) is indicated on each graph. Ages of North

Atlantic Heinrich events are indicated by horizontal lines (labelled H1–H6), which are adjusted to the relevant timescale. Note the various units and grain size fractions for IRD. PS1900 (ref. 29), M23199 and M23071 (ref. 12) are 125–500 μm ; M23385 and NP90-39 are $>500\mu\text{m}$ (ref. 30), PS1730 and PS1919 are $>2\text{mm}$ counted from X-radiographs (5-point moving average) (ref. 4); and PS1297 and PS1294 are $>500\mu\text{m}$ (ref. 31).

hundreds to a few thousands of years (ref. 27). These factors may also have contributed to a lack of coherent millennial-scale oscillations, correlating with Dansgaard–Oeschger temperature variations in the Greenland ice core at Summit²⁸, in the IRD signal in the Nordic seas³ (Fig. 1). Finally, it is clear that both additional dating control on IRD layers and petrographical evidence on the source areas of IRD are important in unravelling what appears an increasingly complex picture of ice-sheet variability¹⁴. □

Received 23 March; accepted 8 June 1999.

1. Heinrich, H. Origin and consequences of cyclic ice rafting in the northeast Atlantic Ocean during the past 130,000 years. *Quat. Res.* **29**, 143–152 (1988).
2. Bond, G. *et al.* Evidence for massive discharges of icebergs into the glacial Northern Atlantic. *Nature* **360**, 245–249 (1992).
3. Fronval, T., Jansen, E., Bloemendal, J. & Johnsen, S. Oceanic evidence for coherent fluctuations in Fennoscandian and Laurentide ice sheets on millennial timescales. *Nature* **374**, 443–446 (1995).
4. Stein, R., Nam, S. I., Grobe, H. & Hubberten, H. in *Late Quaternary Palaeoceanography of the North Atlantic Margins* (eds Andrews, J. T., Austin, W. E. N., Bergsten, H. & Jennings, A. E.) 135–151 (Spec. Publ. 111, Geological Soc., London, 1996).
5. Elliot, M. *et al.* Millennial-scale iceberg discharges in the Irminger Basin during the last glacial period: relationships with the Heinrich events and environmental settings. *Paleoceanography* **13**, 433–446 (1998).

6. MacAyeal, D. R. Binge/purge oscillations of the Laurentide Ice Sheet as a cause of the North Atlantic's Heinrich Events. *Paleoceanography* **8**, 775–784 (1993).
7. Marshall, S. J. & Clarke, G. K. C. A continuum mixture model of ice stream thermo-mechanics in the Laurentide Ice Sheet. 2. Application to the Hudson Strait Ice Stream. *J. Geophys. Res.* **102**, 20615–20637 (1997).
8. Vorren, T. O. *et al.* The Norwegian–Greenland Sea continental margins: morphology and Late Quaternary sedimentary processes and environment. *Quat. Sci. Rev.* **17**, 273–302 (1998).
9. Dowdeswell, J. A. *et al.* Large-scale sedimentation on the glacier-influenced Polar North Atlantic margins: long-range side-scan sonar evidence. *Geophys. Res. Lett.* **23**, 3535–3538 (1996).
10. King, E. L., Sejrup, H. P., Halldason, H., Everhøi, A. & Aarseth, I. Quaternary seismic stratigraphy of the North Sea Fan: glacially-fed gravity flow aprons, hemipelagic sediments, and large submarine slides. *Mar. Geol.* **130**, 293–316 (1996).
11. McManus, J. E., Anderson, R. F., Broecker, W. S., Fleisher, M. Q. & Higgins, S. M. Radiometrically determined sedimentary fluxes in the sub-polar North Atlantic during the last 140,000 years. *Earth Planet. Sci. Lett.* **155**, 29–43 (1998).
12. Baumann, K.-H. *et al.* Reflection of Scandinavian Ice Sheet fluctuations in Norwegian Sea sediments during the past 150,000 years. *Quat. Res.* **43**, 185–197 (1995).
13. Andrews, J. T., Cooper, T. A., Jennings, A. E., Stein, A. B. & Eriksen, H. Late Quaternary iceberg-rafted detritus events on the Denmark Strait–Southeast Greenland continental slope ($\sim 65^\circ\text{N}$): related to North Atlantic Heinrich Events? *Mar. Geol.* **149**, 211–228 (1998).
14. Andrews, J. T. Abrupt changes (Heinrich events) in late Quaternary North Atlantic marine environments: a history and review of data and concept. *J. Quat. Sci.* **13**, 3–16 (1998).
15. Hebbeln, D., Heinrich, R. & Baumann, K. H. Palaeoceanography of the last interglacial–glacial cycle in the Polar North Atlantic. *Quat. Sci. Rev.* **17**, 125–153 (1998).
16. Dowdeswell, J. A., Kenyon, N. H. & Laberg, J. S. The glacier-influenced Scoresby Sund Fan, East Greenland continental margin: evidence from GLORIA and 3.5 kHz records. *Mar. Geol.* **143**, 207–221 (1997).

17. Alley, R. B., Blankenship, D. D., Rooney, S. T. & Bentley, C. R. Sedimentation beneath ice shelves—the view from Ice Stream B. *Mar. Geol.* **85**, 101–120 (1989).
18. Aksu, A. E. & Hiscott, R. N. Shingled Quaternary debris flow lenses on the north-east Newfoundland slope. *Sedimentology* **39**, 193–206 (1992).
19. Hiscott, R. N. & Aksu, A. E. Submarine debris flows and continental slope evolution in front of Quaternary ice sheets, Baffin Bay, Canadian Arctic. *Am. Assoc. Petrol. Geol. Bull.* **78**, 445–460 (1994).
20. Hesse, R., Klauke, L., Ryan, W. B. F. & Piper, D. J. W. Ice-sheet sourced juxtaposed turbidite systems in the Labrador Sea. *Geosci. Can.* **24**, 3–14 (1997).
21. Dowdeswell, J. A. & Siegert, M. J. Ice-sheet numerical modeling and marine geophysical measurements of glacier-derived sedimentation on the Eurasian Arctic continental margins. *Geol. Soc. Am. Bull.* **111**, 1080–1097 (1999).
22. Elverhøi, A. et al. The growth and decay of the Late Weichselian ice sheet in western Svalbard and adjacent areas based on provenance studies of marine sediments. *Quat. Res.* **44**, 303–316 (1995).
23. Fisher, D. A., Reeh, N. & Langley, K. Objective reconstruction of the Late Wisconsinan Laurentide Ice Sheet and the significance of deformable beds. *Geogr. Phys. Quat.* **39**, 229–238 (1985).
24. Hesse, R. et al. Imaging Laurentide Ice Sheet drainage into the deep sea: impact on sediments and bottom water. *GSA Today* **6**, 3–9 (1996).
25. Hesse, R. & Khadabakhsh, S. Depositional facies of late Pleistocene Heinrich events in the Labrador Sea. *Geology* **26**, 103–106 (1998).
26. Dowdeswell, J. A. & Murray, T. in *Glacimarine Environments: Processes and Sediments* (eds Dowdeswell, J. A. & Scourse, J. D.) 121–137 (Spec. Publ. 53, Geological Soc., London, 1990).
27. Johannesson, T., Raymond, C. & Waddington, E. Time-scale for adjustment of glaciers to changes in mass balance. *J. Glaciol.* **35**, 355–369 (1989).
28. Dansgaard, W. et al. Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
29. Baumann, K.-H., Lackschewitz, K. S., Erlenkeuser, H., Henrich, R. & Jünger, B. Late Quaternary calcium carbonate sedimentation and terrigenous input along the East Greenland continental margin. *Mar. Geol.* **114**, 13–36 (1993).
30. Dokken, T. & Hald, M. Rapid climatic shifts during isotope stages 2–4 in the Polar North Atlantic. *Geology* **24**, 599–602 (1996).
31. Hebbeln, D. & Wefer, G. Late Quaternary paleoceanography in the Fram Strait. *Paleoceanography* **12**, 65–78 (1997).

Acknowledgements. This work was supported by the UK NERC and the Marine Science and Technology (ENAM II) programme of the EC. It is a contribution to the ESF PONAM Programme.

Correspondence and requests for materials should be addressed to J.A.D. (e-mail: j.a.dowdeswell@bristol.ac.uk).

Interactions among quantitative traits in the course of sympatric speciation

Alexey S. Kondrashov* & Fyodor A. Kondrashov†

* National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA

† Simon's Rock College, 84 Alford Road, Great Barrington, Massachusetts 01230, USA

Sympatric speciation, the origin of two or more species from a single local population, has almost certainly been involved in formation of several species flocks^{1–4}, and may be fairly common in nature⁵. The most straightforward scenario for sympatric speciation requires disruptive selection favouring two substantially different phenotypes, and consists of the evolution of reproductive isolation between them followed by the elimination of all intermediate phenotypes⁶. Here we use the hypergeometric phenotypic model^{7–10} to show that sympatric speciation is possible even when fitness and mate choice depend on different quantitative traits, so that speciation must involve formation of covariance between these traits. The increase in the number of variable loci affecting fitness facilitates sympatric speciation, whereas the increase in the number of variable loci affecting mate choice has the opposite effect. These predictions may enable more cases of sympatric speciation to be identified.

According to Darwin, competition for diverse resources can lead to disruptive selection and, eventually, to sympatric speciation. "... the more diversified the descendants from any one species become ..., by so much will they be better enabled to seize on many and widely diversified places in the policy of nature, and so be enabled to increase in numbers. ... the competition will generally be most severe between those forms which are most nearly related to each other in habits, constitution, and structure. Hence all the intermediate forms ..., as well as the original parent-species itself,

will generally tend to become extinct" (ref. 11, pp. 112, 121).

Even closely related species usually differ from each other at least at several loci⁵. Thus, multilocus models of sympatric speciation must be studied because single-locus models have very peculiar properties^{6,12}. When a polygenic, quantitative trait is under disruptive selection, indirect selection favours reproductive isolation between the opposite, extreme phenotypes^{7,9,13–15}. If such isolation evolves as a result of fixation of alleles that cause assortative mating based on this trait^{16,17}, the same variable quantitative trait will determine both fitness and mate choice of an individual. Then, according to the data from numerical analysis^{9,10} and individual-based computer simulations^{15,18}, extinction of intermediate phenotypes and sympatric speciation are possible.

Here we study a more complex and perhaps a more realistic case where fitness and mate choice depend on different quantitative traits. We assume that variability in each trait is caused by its own set of independently transmitted loci so that speciation must involve the establishment of a linkage disequilibrium between polymorphic loci controlling fitness and mate choice. This presents an extra obstacle to speciation^{16,17,19}, although some individual-based simulations¹⁵ have suggested that it may still be possible.

In the first model, disruptive selection depends on trait 1 (size) and mate choice depends on trait 2 (colour). Colour is manifested in both females and males and the ability of two individuals to mate depends on the difference between their colours. A typical case of sympatric speciation in a population that was initially almost monomorphic in both traits is presented in Figs 1 and 2. First, frequency-dependent selection, caused by availability of two distinct resources^{9,18}, rapidly establishes polymorphism in size such that individuals with intermediate sizes have low frequencies because they lose out in competition for either resource. Then, disruptive selection slowly increases the magnitude of the covariance between size and colour, owing to the increasing linkage disequilibria between the corresponding loci. Variability of colour also increases slowly, being driven by indirect selection resulting from this covariance. When colour becomes significantly variable, the intertrait covariance increases rapidly and individuals of intermediate sizes and colours disappear rapidly, completing sympatric speciation in the course of relatively few generations. As in the single-locus case¹⁶, the phenotypes of the two new species depend on the sign of the initial intertrait covariance. Thus, these species may consist either of 'small red' and 'large blue' or of 'small blue' and 'large red' individuals.

We can think of this process as the recruitment of colour for providing reproductive isolation between the individuals of the

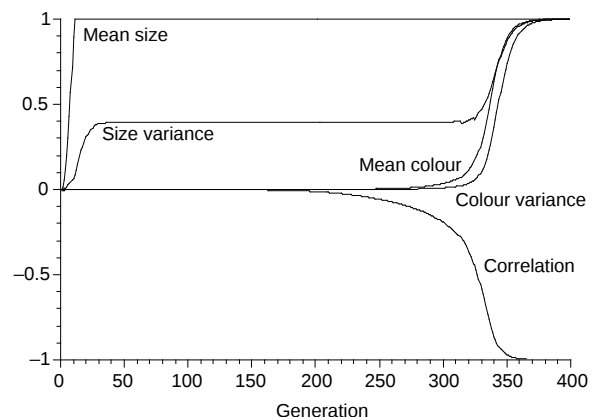


Figure 1 The dynamics of means and variances of size and colour and of the intertrait correlation in the course of sympatric speciation with two traits. The means and variances of size and colour are shown as ratios over their values corresponding to the coexistence of two equally frequent species. Variability of size depends on 16 loci and variability of colour depends on 8 loci. The relative fitness of individuals with intermediate size is 0.15.

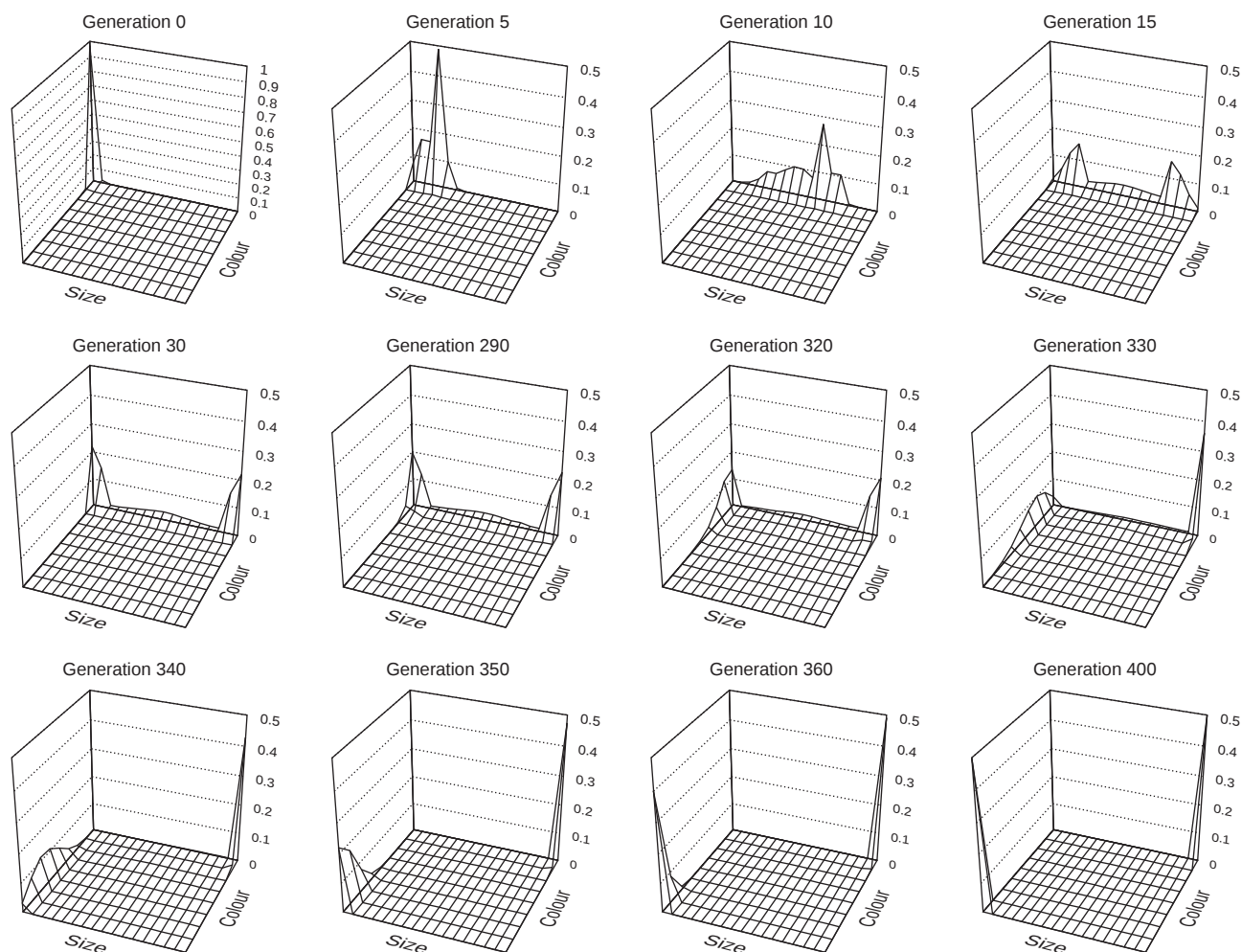


Figure 2 The dynamics of phenotype frequencies in the course of sympatric speciation in the same run as in Fig. 1.

opposite size. A population with zero intertrait covariance can be at equilibrium, but this equilibrium may be locally unstable. If so, recruitment can begin from the increase of even a small variance created by random drift or other factors¹⁶. Similarly, an equilibrium with zero linkage disequilibrium can be unstable in the two-locus selection–recombination model under symmetric selection²⁰.

Table 1 shows the minimal depths of the fitness valley, defined as the ratio of the fitness of individuals with intermediate sizes over that of individuals with extreme sizes, which is necessary for the increase of the initially small covariance and eventual sympatric speciation. When differences at just two haploid loci are enough for reproductive isolation, speciation occurs as a result of assortative mating alone, even without selection¹². When variability of size is due to only two loci, extreme genotypes are frequent even in a panmictic population and speciation does not happen. Generally, the increase in the number of loci affecting size facilitates speciation, and the increase in the number of loci affecting colour has the opposite effect.

In the second model, disruptive selection depends on trait 1 (size), female mate choice depends on trait 2 (preference), and male attractiveness to females depends on trait 3 (colour). The willingness of a female to mate a male depends on the difference between their preference and colour. A typical course of speciation is presented in Fig. 3. The first stage of the process, the establishment of polymorphism in size, is the same as in the first model. In addition, the small positive covariance between preference and colour develops rapidly, owing to assortative mating. Size–preference and size–colour covariances then increase slowly, causing a

slow increase in the variabilities of preference and colour. Eventually, these processes accelerate and all intermediate individuals disappear rapidly. The resulting pair of new species may consist either of ‘small, red-preferring red’ and ‘large, blue-preferring blue’ or of ‘small, blue-preferring blue’ and ‘large, red-preferring red’ individuals, depending on the signs of initial size–preference and size–colour covariances. Table 2 shows the minimal depths of the fitness valley required for the growth of initially small covariances and eventual speciation. If variability in both preference and colour is due to just one haploid locus, no selection is necessary for speciation¹².

Thus, sympatric speciation resulting from disruptive selection and assortative mating is possible even when variability of fitness and mate choice depend on different, independently inherited quantitative traits, provided that disruptive selection is strong and individuals with even similar genotypes make substantially different mate choices. This conclusion is consistent with the results of individual-based computer simulations (ref. 15, and A.S.K. and F.A.K., unpublished results). Naturally, the conditions required for speciation in this case are more stringent than when the same trait determines variability of fitness and mate choice^{15,18}. However, neither catastrophic selection nor complete reproductive isolation resulting from differences at one or two loci are required.

Instead, disruptive selection that is capable of at least doubling the variance in the fitness trait is usually necessary. Even with random mating, such selection establishes bimodal distribution of phenotypes and limits genetic exchange between the opposite extreme phenotypes, owing to low fitness of the offspring from their

Table 1 Maximal relative fitness of intermediate phenotypes that still lead to sympatric speciation in the two-trait model

Number of loci Size	Colour 2	4	8	16	32
2	1.00	Never	Never	Never	Never
4	1.00	0.302	Never	Never	Never
8	1.00	0.377	0.178	0.094	0.033
16	1.00	0.398	0.194	0.169	0.149
32	1.00	0.416	0.194	0.170	0.158

Table 2 Maximal relative fitness of intermediate phenotypes that still lead to sympatric speciation in the three-trait model*

Number of loci Female preference	Male colour	1	2	4	8	16
1	2, 1.00	2, never	2, never	2, never	2, never	2, never
	4, 1.00	4, 0.45	4, 0.02	4, never	4, never	4, never
	8, 1.00	8, 0.42	8, 0.24	8, 0.16	8, 0.10	8, 0.10
	16, 1.00	16, 0.43	16, 0.23	16, 0.19	16, 0.17	16, 0.17
2		2, never	2, never	2, never	2, never	2, never
		4, 0.13	4, never	4, never	4, never	4, never
		8, 0.24	8, 0.23	8, 0.23	8, 0.23	8, 0.23
		16, 0.25	16, 0.23	16, 0.23	16, 0.23	16, 0.23
4			2, never	2, never	2, never	2, never
			4, never	4, never	4, never	4, never
			8, 0.18	8, 0.16	8, 0.15	8, 0.15
			16, 0.19	16, 0.19	16, 0.19	16, 0.19
8				2, never	2, never	2, never
				4, never	4, never	4, never
				8, 0.10	8, 0.08	8, 0.08
				16, 0.17	16, 0.17	16, 0.17
16					2, never	2, never
					4, never	4, never
					8, 0.04	8, 0.04
					16, 0.15	16, 0.15

* The numbers of loci affecting trait 1 are shown within cells. Only the upper triangular part of the table is shown, because the data are symmetric.

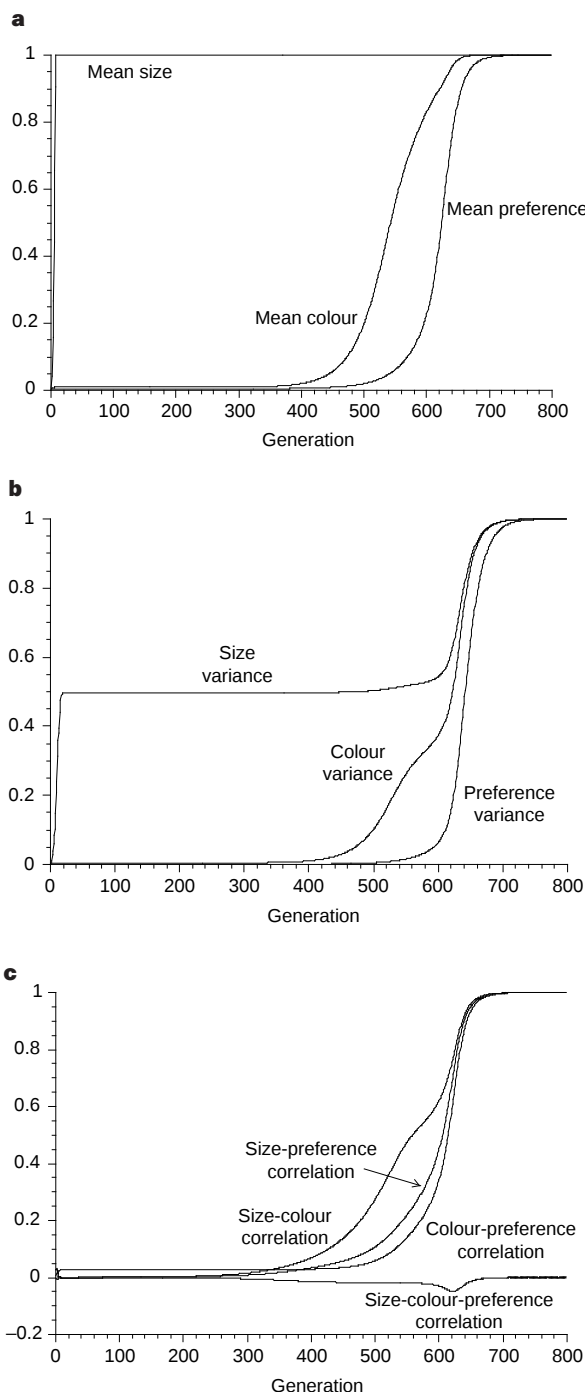


Figure 3 The dynamics of means and variances of size, preference and colour and of the intertrait correlations in the course of sympatric speciation with three traits. **a**, Means; **b**, variances; **c**, intertrait correlations. The means and variances of size, preference and colour are shown as ratios over their values corresponding to the coexistence of two equally frequent species. Variabilities of size, preference and colour depend on 8, 8 and 4 loci, respectively. The relative fitness of individuals with intermediate size is 0.12.

interbreeding^{10,18}. After this, opposite phenotypes may become associated with different values of mate-choice traits, which can lead to the replacement of isolation by selection with true reproductive isolation. Genetic load may stay below 60–70% during the whole course of speciation.

Several scenarios of sympatric speciation attributed to sexual selection^{21–23}, inspired by the apparently sympatric origin of species flocks of African cichlid fish, have been proposed recently. Although the role of sexual selection in the origin of these flocks cannot be ruled out at present, our scenario is more economical and explains all of the data. Although non-random mating due to preference–colour-trait systems^{24–26} currently prevents interbreeding between sympatric species and must have been involved in their origin, such mating does not necessarily lead to sexual selection (that is, to differential reproductive success of different genotypes)²⁷. Assortative mating—that is, non-random mating such that all individuals are equally successful in finding mates²⁷—may be enough. Sympatric species of cichlids often possess different jaw morphology and have very dissimilar diets, using non-overlapping resources²⁴. This suggests that natural as opposed to sexual selection was probably involved in their formation. In addition, because two sympatric species cannot coexist in the same ecological niche, a pair of species that seemed to evolve exclusively as a result of sexual selection would be unstable¹⁸. Indirect selection on mate-choice traits is typically only one order of magnitude weaker than the primary disruptive selection. Thus, variability in mate-choice traits may evolve despite some disadvantage of rare phenotypes as a result of sexual selection.

A pair of sympatric species that have evolved recently according to the scenario described above must have the following properties. First, the two species must be profoundly similar, with interspecific genetic differences confined to loci directly affecting the traits under disruptive selection and responsible for mate choice. Second, interspecific differences in utilization of resources must result from at least several loci. Third, reproductive isolation must be due to differences in not too many loci. Therefore, genetic analysis of morphological, ecological and behavioural differences between sympatric species of a very recent origin^{1–4} can be used to test this scenario.

Methods

Hypergeometric model. If the new species differ from each other at many loci, the number of possible genotypes in the speciating population is too high for the genotype-level analysis, whereas the standard phenotypic approach based on the gaussian approximation clearly cannot describe the splitting of a

population into two. Fortunately, the hypergeometric phenotypic model^{7–9} is applicable under the conditions required for sympatric speciation¹⁰. For a quantitative trait with the possible phenotypes 0, 1, ..., n , this model provides, as a sum of certain hypergeometric functions, the probability $R(i, j, k)$ of two parents with phenotypes i and j producing an offspring with phenotype k (refs 8–10).

Implementation of the model. We considered haploid individuals with the phenotype in their m -th trait determined by the number of alleles 1 at the n_m corresponding loci, each with alleles 0 and 1. The dynamics with $n_m/2$ diploid loci are very similar¹⁰. The generations were discrete and the life cycle consisted of selection, mating and reproduction.

In the two-trait model, the frequency of (i_1, i_2) individuals of the i_1 -th phenotype in trait 1 and i_2 -th phenotype in trait 2 before selection was $p(i_1, i_2)$. After selection, this frequency becomes $p'(i_1, i_2) = w(i_1, i_2)p(i_1, i_2)/W$, where $w(i_1, i_2)$ is the fitness function and W is the mean fitness. An individual mated no more than once. Individuals paired randomly and mating of a pair occurred with the probability $M(d)$, where d is the ratio of the difference between their phenotypes in trait 2 expressed as a proportion of n_2 . All unmated individuals paired again, and the process continued until less than 10^{-10} of the population remained unmated. In this way, $A(i_1, i_2, j_1, j_2)$, that is, the frequency of mating between (i_1, i_2) and (j_1, j_2) individuals, was calculated. During reproduction a pair of such parents produced an offspring (k_1, k_2) with the probability $R(i_1, j_1, k_1)R(i_2, j_2, k_2)$, where R_m describes the transmission of phenotypes in the m -th trait.

The three-trait model was analogous, but M depended on $d = |i_2/n_2 - j_2/n_2|$, where i_2 and j_2 were phenotypes of the first (female) and second (male) potential partners in traits 2 and 3, respectively. THINK C programs are available on request.

Parameters. During selection, 10% of individuals with the highest and lowest values of their phenotypes in trait 1 had fitness 1.0, and the fitness of the rest was $D < 1.0$. Such selection causes sympatric speciation most efficiently¹⁸.

In both models, $M(d) = 1 - d$, that is, the degree of reproductive isolation, grew linearly with the difference between the potential mates, but only the individuals that were maximally different were completely isolated. If such individuals were isolated but all others mated freely, speciation never occurred.

In most runs, the initial population consisted of 99.99% of individuals with phenotype 0 in all two or three traits. The rest of the population had phenotypes no more than 1 in every trait and the covariances between the traits were very small.

Received 20 January; accepted 10 May 1999.

- Mina, M. V., Mironovsky, A. N. & Dgebuadze, Y. Y. Lake Tana large barbs: phenetics, growth and diversification. *J. Fish Biol.* **48**, 383–404 (1996).
- Berri, P. & Valiushok, D. Genetic divergence among morphotypes of Lake Tana (Ethiopia) barbs. *Biol. J. Linn. Soc.* **64**, 369–384 (1998).
- Schlieven, U. K., Tautz, D. & Paabo, S. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* **368**, 629–632 (1994).
- Johnson, T. C. et al. Late pleistocene desiccation of Lake Victoria and rapid evolution of cichlid fishes. *Science* **273**, 1091–1093 (1996).
- Ritchie, M. G. & Phillips, S. D. F. in *Endless Forms: Species and Speciation* (eds Howard, D. J. & Berlocher, S. H.) 291–308 (Oxford Univ. Press, New York, 1998).
- Kondrashov, A. S. & Mina, M. V. Sympatric speciation: when is it possible? *Biol. J. Linn. Soc.* **27**, 201–223 (1986).
- Kondrashov, A. S. On the intensity of selection for reproductive isolation at the beginnings of sympatric speciation. *Genetika* **20**, 408–415 (1984).
- Barton, N. H. On the spread of new gene combinations in the third phase of Wright's shifting-balance. *Evolution* **46**, 551–557 (1992).
- Doebeli, M. A quantitative genetic competition model for sympatric speciation. *J. Evol. Biol.* **9**, 893–909 (1996).
- Shpak, M. & Kondrashov, A. S. Applicability of the hypergeometric phenotypic model to haploid and diploid populations. *Evolution* **53**, 600–604 (1999).
- Darwin, C. *The Origin of Species by Means of Natural Selection* (Murray, London, 1859).
- Kondrashov, A. S. & Shpak, M. On the origin of species by means of assortative mating. *Proc. R. Soc. Lond. B* **265**, 2273–2278 (1998).
- Wallace, A. R. *Darwinism* (Macmillan, London, 1889).
- Rice, W. R. Disruptive selection of habitat preference and the evolution of reproductive isolation: a simulation study. *Evolution* **38**, 1251–1260 (1984).
- Kondrashov, A. S. Multilocus model of sympatric speciation. III. Computer simulations. *Theor. Popul. Biol.* **29**, 1–15 (1986).
- Maynard Smith, J. Sympatric speciation. *Am. Nat.* **100**, 637–650 (1966).
- Rice, W. R. & Hostert, E. E. Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* **47**, 1637–1653 (1993).
- Kondrashov, A. S., Yampolsky, L. Yu. & Shabalina, S. A. in *Endless Forms: Species and Speciation* (eds Howard, D. J. & Berlocher, S. H.) 90–98 (Oxford Univ. Press, New York, 1998).
- Felsenstein, J. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**, 124–138 (1981).
- Karlin, S. General two-locus selection models: some objectives, results and interpretations. *Theor. Popul. Biol.* **7**, 364–398 (1975).

- Turner, G. F. & Burrows, M. T. A model of sympatric speciation by sexual selection. *Proc. R. Soc. Lond. B* **260**, 287–292 (1995).
- Galis, F. & Metz, J. A. J. Why are there so many cichlid species? *Trends Ecol. Evol.* **13**, 1–2 (1998).
- van Doorn, G. S., Noest, A. J. & Hogeweg, P. Sympatric speciation and extinction driven by environment dependent sexual selection. *Proc. R. Soc. Lond. B* **265**, 1915–1919 (1998).
- Axelrod, H. R. *The Most Complete Colored Lexicon of Cichlids* (TFH Publications, Neptune City, 1996).
- Seehausen, O., van Alphen, J. J. M. & Witte, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808–1811 (1997).
- Seehausen, O. & van Alphen, J. J. M. The effect of male coloration on female mate choice in closely related Lake Victoria cichlids (*Haplochromis nyererei* complex). *Behav. Ecol. Sociobiol.* **42**, 1–8 (1998).
- Lewontin, R. C., Kirk, D. & Crow, J. F. Selective mating, assortative mating and inbreeding: definitions and implications. *Eugen. Quart.* **15**, 141–143 (1966).

Acknowledgements. This study was supported by a grant from the NSF.

Correspondence and requests for materials should be addressed to F.A.K. (e-mail: fedya@simons-rock.edu).

On the origin of species by sympatric speciation

Ulf Dieckmann & Michael Doebeli

Adaptive Dynamics Network, International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

Zoology Institute, University of Basel, Rheinsprung 9, CH-4051 Basel, Switzerland

Understanding speciation is a fundamental biological problem. It is believed that many species originated through allopatric divergence, where new species arise from geographically isolated populations of the same ancestral species^{1–3}. In contrast, the possibility of sympatric speciation (in which new species arise without geographical isolation) has often been dismissed, partly because of theoretical difficulties^{2,3}. Most previous models analysing sympatric speciation concentrated on particular aspects of the problem while neglecting others^{4–10}. Here we present a model that integrates a novel combination of different features and show that sympatric speciation is a likely outcome of competition for resources. We use multilocus genetics to describe sexual reproduction in an individual-based model, and we consider the evolution of assortative mating (where individuals mate preferentially with like individuals) depending either on an ecological character affecting resource use or on a selectively neutral marker trait. In both cases, evolution of assortative mating often leads to reproductive isolation between ecologically diverging subpopulations. When assortative mating depends on a marker trait, and is therefore not directly linked to resource competition, speciation occurs when genetic drift breaks the linkage equilibrium between the marker and the ecological trait. Our theory conforms well with mounting empirical evidence for the sympatric origin of many species^{10–18}.

The theory of adaptive dynamics^{19–22} is a general framework for studying phenotypic evolution driven by ecological interactions. One of the phenomena unravelled by adaptive dynamics is evolutionary branching, during which directional selection drives a monomorphic population to a phenotype where ecological interactions induce disruptive selection and a subsequent split into two coexisting phenotypic clusters (Fig. 1a). Evolutionary branching explains the dynamic emergence and perpetuity of disruptive selection and serves as a unifying concept for understanding the evolution of polymorphisms. It is found in a wide range of models of asexual populations (see refs 22 and 23 for examples). Here we show that evolutionary branching also occurs in sexual populations and thus leads to a general theory for sympatric speciation.

We start from assumptions that are likely to be satisfied in many natural populations. Individuals vary in a quantitative character x determining resource use, as for example when beak size in birds determines the size of seeds consumed. Populations consisting of individuals of a given trait value x have density-dependent logistic growth with carrying capacity $K(x)$. We assume that the resource

distribution $K(x)$ is unimodal and varies according to a gaussian function $N(x_0, \sigma_K)$, with the maximum at an intermediate phenotype x_0 and variance σ_K . In polymorphic populations consisting of individuals with different trait values, dissimilar individuals interact only weakly, as, for example, when birds with different beak sizes eat different types of seed. That is, competition is not only density- but also frequency-dependent, and rare phenotypes experience less competition than common phenotypes. Specifically, we assume that the strength of competition between individuals declines

with phenotypic distance according to a gaussian function $N(0, \sigma_C)$, with a maximum at zero and variance σ_C^2 .

These assumptions are integrated into an asexual individual-based model in which each individual is characterized by its trait value x . Individuals give birth at a constant rate and die at a rate that is determined by frequency- and density-dependent competition (see Methods). Evolutionary dynamics occur because offspring phenotypes may deviate slightly from parent phenotypes. The quantitative character first evolves to the value x_0 with maximal carrying capacity. After that, two things can happen: either x_0 is evolutionarily stable and evolution comes to a halt at x_0 , or x_0 is actually a fitness minimum and can be invaded by all nearby phenotypes^{19,21,22}. In the latter case, evolutionary branching occurs

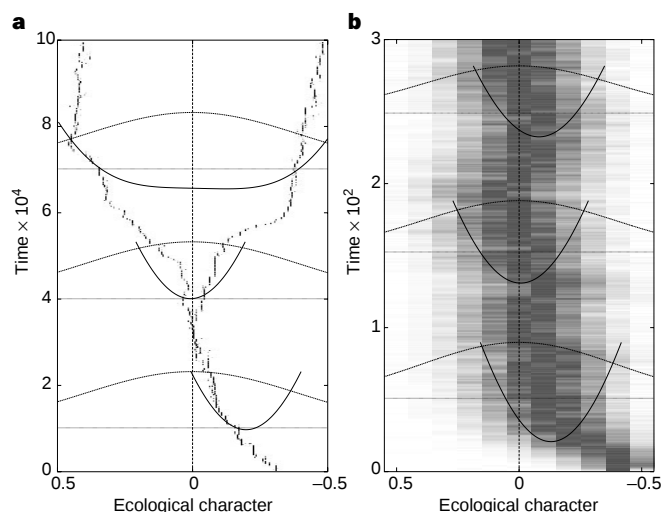


Figure 1 Convergence to disruptive selection **a**, Evolutionary branching in the individual-based asexual model: at the branching point $x_0 = 0$, the population splits into two morphs. Three insets show fitness functions (continuous curves) generated by the ecological interactions at different points in time (indicated by horizontal dotted lines). Selection changes from directional to disruptive when evolution reaches x_0 . The resource distribution $K(x)$ has its maximum at x_0 and is shown for comparison (dashed curve). **b**, As in **a**, but with multilocus genetics for the ecological character and random mating. Shading represents phenotype distributions (5 diploid and diallelic loci result in 11 possible phenotypes). Despite disruptive selection at the branching point (see insets), branching does not occur.

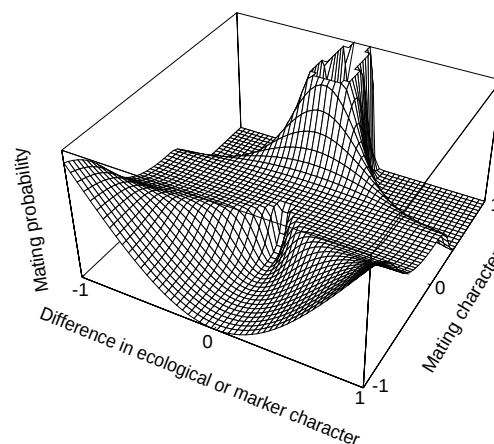


Figure 2 Mating probabilities as determined by mating character and difference in ecological or marker character between mates. The mating character m is scaled to vary between -1 (all $-$ alleles) and $+1$ (all $+$ alleles). Mating probabilities vary with differences in either ecological or marker character, depending on the scenario. If the mating character in the focal individual is close to $+1$, it has a high probability of mating with similar individuals. If its mating character is close to -1 , it is more likely to mate with dissimilar individuals. Intermediate mating characters (close to 0) correspond to random mating.

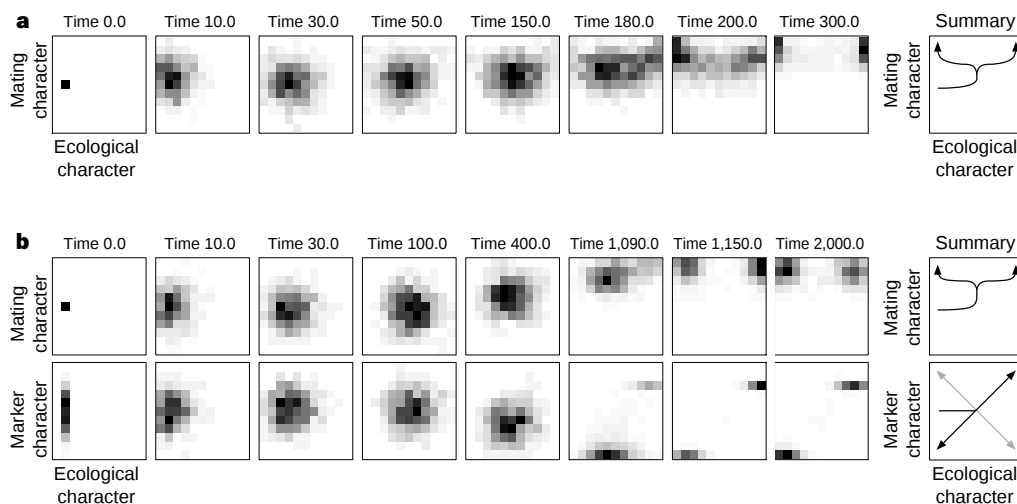


Figure 3 Evolutionary branching in sexual populations. **a**, First scenario: mating probabilities (vertical axes) depend on the ecological character (horizontal axes), which first evolves to intermediate values (50 generations). Then the mean mating character increases to positive values (180 generations) and induces a bimodal split in the ecological character (200 generations). **b**, Second scenario: mating probabilities (vertical axes in upper panels) depend on a marker trait (vertical axes in lower panels). The ecological trait (horizontal axes in all panels) first evolves to intermediate values (100 generations). Owing to temporary

correlations between marker and ecological trait, assortative mating increases, which in turn magnifies these correlations (generations 400–1,090). This positive feedback eventually leads to speciation (1,150 generations). In **b** branching typically takes longer than in **a**. The summary panels depict the evolution of mean character values schematically. Grey arrows in the bottom summary panel show an alternative, equally likely, evolution of linkage disequilibrium between ecological and marker character.

(Fig. 1a). This happens for $\sigma_C < \sigma_K$, that is, if the curvature of the carrying capacity at its maximum is less than that of the competition function. Then the advantage of deviating from the crowded optimal phenotype x_0 more than compensates for the disadvantage of a lower carrying capacity.

Sexual reproduction is incorporated by assuming that character values are determined by many additive, diploid loci with two alleles, + and -, and are proportional to the number of 'plus' alleles. Offspring inherit maternal and paternal alleles at each locus independently (free recombination). As in the asexual case, the sexual population evolves to a mean phenotype x_0 . If mating is random, however, evolutionary branching does not occur for any values of σ_K and σ_C : the split into two distinct phenotypic morphs is prevented by the continual generation of intermediate phenotypes through recombination (Fig. 1b). Thus, in sexual populations, non-random mating is a prerequisite for evolutionary branching²⁴.

To model the evolution of assortative mating we assume that individuals express an additional quantitative character that determines mating probabilities according to two scenarios. In the first, mating probabilities are based on similarity in the ecological character, and in the second they are based on similarity in a third, ecologically neutral 'marker' trait (see Methods). Mating

character and marker trait are also determined by many additive diallelic loci. Individuals with an intermediate mating character mate randomly. Individuals carrying mostly 'minus' alleles at the mating loci mate disassortatively, and hence are more likely to mate with individuals with very different ecological or marker phenotypes. Individuals carrying mostly plus alleles at the mating loci mate assortatively: the probability of mating increases with phenotypic similarity to the partner (Fig. 2).

Figure 3a shows the evolutionary dynamics of an initially randomly mating population when mating probabilities depend on the ecological character. While this character evolves to x_0 , the mating character initially changes only slowly, but it picks up speed and evolves towards positive assortativeness when the ecological character reaches x_0 . Once assortativeness is strong enough, the population splits into two ecologically different morphs which eventually are almost completely reproductively isolated. These results confirm and extend those of ref. 24 and occur because, near the dynamically emerging fitness minimum at x_0 , selection favours mechanisms that allow for a split in the phenotype distribution and hence for a departure from the fitness minimum. Assortative mating is such a mechanism, because it prevents the generation of intermediate offspring phenotypes from extreme parent phenotypes. Parameter requirements for evolutionary branching in sexual populations appear to be only slightly more restrictive than in the asexual case (Fig. 4).

When assortative mating depends on the ecological character, speciation is not hindered by recombination between mating loci and ecological loci. However, when mating depends on an ecologically neutral marker trait, a linkage disequilibrium between marker loci and ecological loci, leading to a correlation between marker trait and ecological character, is required for the evolution of assortative mating and for speciation. Classical, deterministic models (such as Felsenstein's 'two-allele' models⁶) predict that such linkage disequilibria are unlikely because of recombination between ecological and marker loci^{3,6}. In our individual-based model, however, genetic drift due to stochastic demographic effects readily leads to speciation despite the opposing force of recombination. Figure 3b shows the adaptive dynamics when mating probabilities depend on a neutral marker trait. Genetic drift temporarily results in small and localized linkage disequilibria between some marker loci and some ecological loci. Positive and negative correlations both select for assortative mating, which in turn magnifies the local disequilibria into a global linkage disequilibrium between marker and ecological trait. This feedback eventually induces the sympatric split into reproductively isolated phenotypic clusters. Thus, stochastic fluctuations in finite populations can spontaneously break the symmetry of linkage equilibria seen in determi-

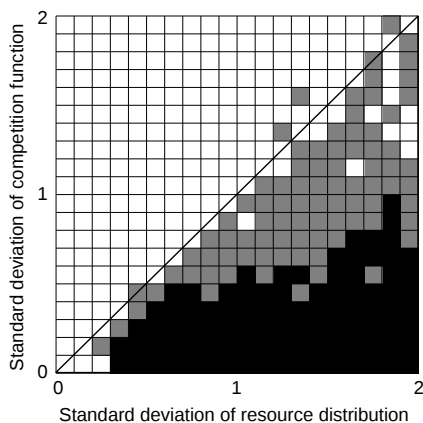


Figure 4 Combinations of standard deviations σ_K and σ_C of the resource distribution $K(x)$ and competition function $C(x)$, respectively, that allow for evolutionary branching. Analytical results are available for the asexual model (see Methods) and predict branching for $\sigma_C < \sigma_K$, that is, below the diagonal. Conditions for branching in sexual populations (within 20,000 generations) are shown in grey when mating probabilities depend on the ecological character and in black when they depend on a marker trait.

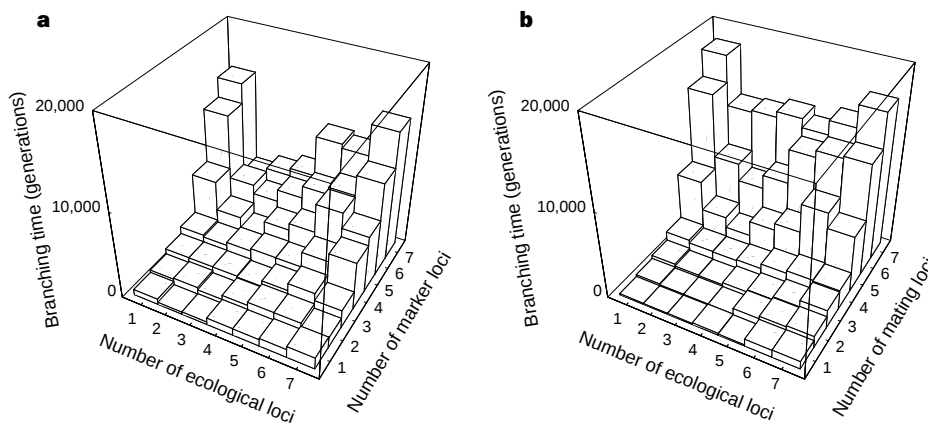


Figure 5 Average waiting times for evolutionary branching with different numbers of loci when assortative mating depends on a marker trait. **a**, Variable numbers of loci for marker and ecological trait with a fixed number of loci (5) for assortative

mating. **b**, Variable numbers of loci for assortative mating and ecological trait with a fixed number of loci (5) for the marker trait. Other parameters are as in Fig. 3; each column represents the average waiting time from 60 simulation runs.

nistic models. Recombination between marker loci and ecological loci implies that parameter requirements for evolutionary branching are more restrictive when mate choice is based on a neutral marker than when it is based on the ecological trait (Fig. 4).

The effects of stochasticity on evolutionary branching are further illustrated by varying the number of loci determining the quantitative characters (Fig. 5). Evolutionary branching is more likely when there are fewer loci, because then the phenotypic effects of genetic drift are larger (an exception occurs with only one ecological locus: with only three phenotypes, sufficiently strong fluctuations arise more rarely). Branching triggered by drift becomes less likely in very large populations where stochastic effects become small.

Our results extend and contrast previous insights^{6,8,9,24–26} by showing that competition for unimodal resources can initiate sympatric speciation even if assortative mating depends on an ecologically neutral marker trait. The results are robust against changes in the models such as varying numbers of loci (Fig. 5), assuming different mutation rates per locus, different relations between the number of plus alleles on the mate choice loci and the degree of assortativeness (see Methods), and different functions for the carrying capacities, $K(x)$, and for the strength of competition, $C(x)$, while maintaining their qualitative characteristics. Evidence is accumulating that ecology is important for speciation^{18,27,28}, and our theory may provide an integrative framework for understanding otherwise puzzling evidence for monophyletic origins of many sympatric species, including cichlids^{11,12}, sticklebacks^{13,16,27}, snails¹⁴, giant senecios¹⁵ and anolis lizards¹⁷. In all these cases, it is likely that frequency-dependent mechanisms are important determinants of the species' ecologies. Therefore, assortative mating based on ecologically important traits such as body size (as in sticklebacks²⁹) or on marker traits that co-vary with ecological traits (such as coloration or breeding behaviour in cichlids³⁰) could have led to the formation of new species in accordance with the theory presented here. We expect our theory to work best in relatively recently colonized habitats, in which sympatric divergence is not strongly opposed by competition from other species already present. In fact, a striking example of incipient sympatric speciation due to ecological interactions in a new habitat has recently been documented in a pair of cichlid morphs (U. K. Schliewen *et al.*, submitted), in which restricted gene flow has evolved through size-assortative mating. The mechanisms of speciation are rarely as clear as in this example, but our theoretical evidence generally suggests a prominent role for ecologically driven speciation in sympatry. □

Methods

The deterministic dynamics of a resident population of phenotype x are

$$\frac{dN(x, t)}{dt} = r \cdot N(x, t) \cdot \left[1 - \frac{N(x, t)}{K(x)} \right]$$

where $N(x, t)$ is the population size at time t . The carrying capacity, $K(x) = K_0 \cdot \exp(-\frac{(x-x_0)^2}{2\sigma_K^2})$, is the stable equilibrium. When a rare mutant y appears in a resident x at carrying capacity $K(x)$, it competes with the discounted density $C(x-y) \cdot K(x)$, where $C(x-y) = \exp(-\frac{(x-y)^2}{2\sigma_C^2})$ describes the strength of competition between phenotypes. Therefore, the per capita growth rate $s(y, x)$ of the rare mutant y is $r \cdot [1 - \frac{C(x-y) \cdot K(x)}{K(y)}]$. The derivative $\frac{\partial s(y, x)}{\partial y} \Big|_{y=x} = r \cdot \frac{K'(x)}{K(x)}$ of $s(y, x)$ with respect to the mutant y and evaluated at the resident x is positive for $x < x_0$ and negative for $x > x_0$. Therefore, x_0 is an attractor for the adaptive dynamics^{19,21,22}. In addition, if $s(y, x_0)$ has a minimum at $y = x_0$, then x_0 is a branching point^{19,21,22}. This happens if and only if $\sigma_C < \sigma_K$.

These analytical predictions are confirmed by the individual-based asexual model, in which individuals are assigned a phenotype x , give birth at a rate r and die at a rate $\frac{r}{K(x)} \cdot \sum_y N(y, t) \cdot C(x-y)$, where the sum weighs all individuals by their competitive impact on x . Offspring have the same phenotype as their parent, except when a mutation occurs (at rate 0.001), in which case their phenotype is chosen from a normal distribution $N(x, \frac{1}{2\sigma_K^2})$, where x is the parent phenotype.

In sexual populations, birth and death rates are calculated similarly. Individuals are assigned up to three diploid genotypes with five diallelic loci

each (variation in loci number is analysed in Fig. 5). The first set of loci determines the ecological character x , the second set determines mating probabilities and the third encodes the marker trait. The mating character m is given by the difference between the number of + and - alleles divided by the total number of alleles. If assortative mating depends on the ecological trait, then, for $m > 0$, mating probabilities fall off with a difference in the ecological trait according to a gaussian function $N(x, \sigma_d)$ with mean equal to the focal individual's ecological trait and variance $\sigma_d = \frac{1}{2\sigma_m^2}$. If $m = 0$, the focal individual mates randomly. If $m < 0$, then mating probabilities increase with ecological difference according to the function $1 - N(x, \sigma_d)$, where $\sigma_d = \frac{1}{m^2}$ (Fig. 2). If assortative mating depends on the marker trait, then the third set of loci replaces the ecological trait in determining mating probabilities, which then depend on similarity in the marker trait. To avoid a bias against marginal phenotypes in the population, mating probabilities are normalized, so that the sum of mating probabilities over all potential partners is 1 for all phenotypes. A 50:50 sex ratio is assumed at all times. At each locus, one offspring allele is chosen randomly from the two maternal alleles and the other from the two paternal alleles at this locus. With a small probability (0.001), a mutation occurs in the inherited alleles and reverses their value. Other parameter values used for the figures are $r = 1$, $K_0 = 500$, $\sigma_K = 1$ and $\sigma_C = 0.4$ (variation in the last two parameters is analysed in Fig. 4).

Received 28 January; accepted 1 June 1999.

- Mayr, E. *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, Massachusetts, 1963).
- Coyne, J. A. Genetics and speciation. *Nature* **355**, 511–515 (1992).
- Rice, W. R. & Hostert, E. E. Laboratory experiments on speciation—what have we learned in 40 years. *Evolution* **47**, 1637–1653 (1993).
- Maynard Smith, J. Sympatric speciation. *Am. Nat.* **100**, 637–650 (1966).
- Rosenzweig, M. L. Competitive speciation. *Biol. J. Linn. Soc. (Lond.)* **10**, 275–289 (1978).
- Felsenstein, J. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**, 124–138 (1981).
- Seger, J. in *Evolution. Essays in Honour of John Maynard Smith* (eds Greenwood, P. J., Harvey, P. H. & Slatkin, M.) (Cambridge Univ. Press, Cambridge, 1985).
- Kondrashov, A. S. Multilocus model of sympatric speciation III. computer simulations. *Theor. Pop. Biol.* **29**, 1–15 (1986).
- Johnson, P. A., Hoppenstaedt, F. C., Smith, J. J. & Bush, G. L. Conditions for sympatric speciation: a diploid model incorporating habitat fidelity and non-habit assortative mating. *Evol. Ecol.* **10**, 187–205 (1996).
- Bush, G. L. Sympatric speciation in animals—new wine in old bottles. *Trends Ecol. Evol.* **9**, 285–288 (1994).
- Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **347**, 550–553 (1990).
- Schliewen, U. K., Tautz, D. & Pääbo, S. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* **368**, 629–633 (1994).
- Schluter, D. Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266**, 798–801 (1994).
- Johannesson, K., Rolan-Alvarez, E. & Ekendahl, A. Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. *Evolution* **49**, 1180–1190 (1995).
- Knox, E. B. & Palmer, J. D. Chloroplast DNA variation and the recent radiation of giant senecios (Asteraceae) on the tall mountains of Eastern Africa. *Proc. Natl Acad. Sci. USA* **92**, 10349–10353 (1995).
- Taylor, E. B. & McPhail, J. D. Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (*Gasterosteus*): insights from mitochondrial DNA. *Biol. J. Linn. Soc.* **66**, 271–291 (1999).
- Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K. & Rodriguez-Schettino, L. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**, 2115–2118 (1998).
- Orr, M. R. & Smith, T. B. Ecology and speciation. *Trends Ecol. Evol.* **13**, 502–506 (1998).
- Metz, J. A. J., Geritz, S. A. H., Meszéna, G., Jacobs, F. J. A. & van Heerwaarden, J. S. in *Stochastic and Spatial Structures of Dynamical Systems* (eds van Strien, S. J. & Verdun Lunel, S. M.) 183–231 (North Holland, Amsterdam, 1996).
- Dieckmann, U. & Law, R. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* **34**, 579–612 (1996).
- Dieckmann, U. Can adaptive dynamics invade? *Trends Ecol. Evol.* **12**, 128–131 (1997).
- Geritz, S. A. H., Kisdi, E., Meszéna, G. & Metz, J. A. J. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57 (1998).
- Doebeli, M. & Ruxton, G. D. Evolution of dispersal rates in metapopulation models: branching and cyclic dynamics in phenotype space. *Evolution* **51**, 1730–1741 (1997).
- Doebeli, M. A quantitative genetic competition model for sympatric speciation. *J. Evol. Biol.* **9**, 893–909 (1996).
- Turner, G. F. & Burrows, M. T. A model of sympatric speciation by sexual selection. *Proc. R. Soc. Lond. B* **260**, 287–292 (1995).
- Kondrashov, A. S. & Shpak, M. On the origin of species by means of assortative mating. *Proc. R. Soc. Lond. B* **265**, 2273–2278 (1998).
- Schluter, D. & McPhail, J. D. Character displacement and replicate adaptive radiation. *Trends Ecol. Evol.* **8**, 197–200 (1993).
- Schluter, D. & Nagel, L. M. Parallel speciation by natural selection. *Am. Nat.* **146**, 292–301 (1995).
- Nagel, L. M. & Schluter, D. Body size, natural selection, and speciation in sticklebacks. *Evolution* **52**, 209–218 (1998).
- Seehausen, O., van Alphen, J. J. M. & Witte, F. Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* **277**, 1808–1811 (1997).

Acknowledgements. We thank H. Metz, R. Law, M. Heino, S. Stearns, J. Maynard Smith and N. Knowlton for helpful comments, and U. Schliewen, K. Rasmussen and D. Tautz for discussions. The order of authors is alphabetical.

Correspondence and requests for materials should be addressed to U.D. (e-mail: dieckman@iiasa.ac.at).

Environmental determination of a sexually selected trait

Simon C. Griffith^{††}, Ian P. F. Owens[§] & Terry Burke[‡]

^{*} Department of Biology, University of Leicester, Leicester LE1 7RH, UK

[†] Department of Animal Ecology, Uppsala University, S-752 36 Uppsala, Sweden

[§] Department of Zoology and Entomology, University of Queensland, St Lucia, Brisbane 4072, Australia

Models of sexual selection usually assume that variation in the expression of sexual ornaments is determined largely by genetic, rather than environmental, factors¹. However, empirical support for this assumption comes from studies of species with little parental care^{1,2}, in which the influence of environmental factors may be limited^{3,4}, and from studies of just two species^{5,6} with parental care, in both of which heritability estimates vary hugely between years or populations^{7,8}. In the remaining studies of species with parental care, it is not known whether resemblance in sexual ornamentation between relatives was due to shared genes or shared patterns of care^{3,4,9–15}. Here we use cross-fostering experiments in house sparrows, *Passer domesticus*, to examine the relative roles of these effects. We demonstrate that, although sons resemble their fathers with respect to sexual ornamentation, this resemblance is mainly due to post-hatching environmental effects rather than shared genes. We also show that sons hatching early in the year have the largest ornaments. These results support models that emphasize the importance of environmental sources of variation^{16–18}, such as direct paternal effects^{4,19,20}, on the expression of sexual ornaments, and agree with the general observation that sexually selected traits tend to be condition dependent¹. We urge the incorporation of gene–environment interactions into future models of sexual selection.

We studied the population of house sparrows on Lundy, an island in the Bristol Channel, UK, between 1995 and 1996. During this period there were about 100 breeding adults on the island, almost all of which were individually marked for identification and nested in artificial nest-boxes. Although the population is ecologically isolated, being approximately 20 km offshore, there is immigration from the mainland (on average, 0.75 individuals per year; or 1.5 breeding individuals per generation²¹), and molecular evidence from highly polymorphic microsatellite loci shows that the population is neither inbred nor genetically poor compared with mainland populations²¹. We used cross-fostering experiments to study the inheritance of the black throat patch, or 'badge', of male sparrows in this population. This badge is expressed only in males, and males with large badges have been shown to be behaviourally dominant^{22–24} and preferred by females as social and extra-pair copulation partners^{25,26}. Males grow the badge only once a year, soon after the end of the breeding season. On average, the badges of first-year males are slightly smaller (length, 32.11 ± 0.39 mm (mean $\pm$ s.e.); $n = 35$) than those of second-year males (34.02 ± 0.52 ; $n = 20$) whose badges are approximately the same size as older males (34.22 ± 0.44 ; $n = 29$) (one-way analysis of variance, $F_{2,83} = 7.44$, $P < 0.01$). Nevertheless, the badge grown by first-year males when they are a few months old is the badge they will have during their first reproductive attempt. During cross-fostering experiments, offspring were swapped between pairs of nests so that each offspring had both biological and foster parents. The relative similarity between sons and their biological father versus sons and their foster father provides an index of the relative importance of genetic and environmental determinants of badge size^{3,5–7}. Allele sharing at

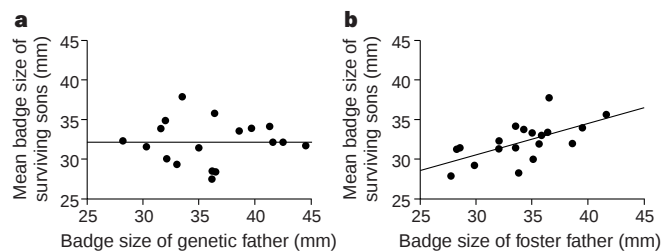


Figure 1 Correlations between the badge size of male offspring and the badge size of genetic and foster fathers. **a**, Genetic father ($r^2 = 0.001$, $n = 18$, $P > 0.5$, $y = -0.003x + 32.17$) and **b**, foster father ($r^2 = 0.38$, $n = 19$, $P = 0.005$, $y = 0.4x + 18.76$), where x is the father's badge size, y is the offspring's badge size and r^2 is the coefficient of determination. Narrow sense heritability was estimated as twice the slope of the regression³. Young sparrows moult into the badge that they will have during their first breeding season, several months after fledging. The analysis was therefore restricted to those offspring who survived this period and were recaptured. Mean offspring badge size was used in cases ($n = 3$) where a male had more than one surviving son.

microsatellite loci was used to ensure that paternity had been assigned accurately²¹.

Contrary to the predictions of good-genes models of sexual selection, the badge sizes of sons did not resemble the badge sizes of their biological fathers (Fig. 1a; slope = -0.003 , $h^2 = 0.006 \pm 0.15$, $P > 0.5$) but did significantly resemble the badge size of their foster fathers (Fig. 1b; slope = 0.40 ± 0.12 , $P = 0.005$). Furthermore, the differences between biological- and foster-father effects with respect to badge size did not appear to be caused by the differential survival of sons before measurement, as there was no significant correlation between offspring survival and the badge size of either their biological or foster father (logistic regression of offspring survival to six months on each type of father's badge size: $r^2 < 0.001$, $n = 49$, d.f. = 2, $P > 0.99$).

These observations indicate that the previously reported inheritance of badge size in sparrows¹⁰ is largely due to non-genetic sources, supporting the idea that badge size may, in part at least, be determined by environmental factors²⁷. The previously reported heritability estimate for sparrow badge size was 0.60 ± 0.23 (ref. 10), which is not significantly different from the 'heritability' of 0.80 ± 0.24 estimate from the regression of sons on foster fathers. The 95% confidence limits around our new estimate of genetic heritability are from -0.32 to 0.33 . Thus, although we cannot rule out the possibility that genetic sources of variation are sometimes important in determining variation in this trait, our results strongly support the idea that the expression of badge size in house sparrows is strongly influenced by environmental factors. This conclusion corroborates previous work indicating that badge size in house sparrows is condition dependent²⁷, and our finding that it is a highly phenotypically plastic trait (Spearman rank correlation between the badge size in 1995 versus the badge size in 1996 of the same individual, $r_s = 0.25$, $n = 31$, $P > 0.10$).

To investigate potential environmental sources of variation, we fitted a multiple-regression model to variation in badge size among sons. As well as the badge sizes of foster fathers, we included independent variables relating to the size of the sons, their date of hatching, the number of siblings in the nest and the rate at which they were fed by each parent. Given our small sample size, a stepwise model was used to minimize the chances of spurious variables being included in the final model. Of the additional independent variables, only hatching date was significantly associated with the badge size of sons, with sons hatching early having larger badges. This result is consistent with previous studies showing that hatching date has a strong influence on offspring quality⁴. The multiple-regression model including son's hatching date and the foster-father's badge size explained more than 70% of the variation in badge size amongst

[‡] Present addresses: Department of Evolutionary Biology, Uppsala University, S-752 36 Uppsala, Sweden (S.C.G.); Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK (T.B.).

Table 1 Stepwise multiple-regression model of son's badge size against potential environmental sources of variation

Independent variables in final model	F-ratio	Coefficient	P
Badge size of foster father	20.61	0.41	<0.001
Date of hatching of son	12.07	-0.07	<0.005
Independent variables not in final model	F-ratio	Coefficient	P
Badge size of genetic father	0.18	-0.07	>0.50
Tarsus length of son	0.08	-0.30	>0.50
Residual mass of son	1.23	-0.28	>0.10
Number of eggs laid in foster clutch	0.20	-0.73	>0.50
Number of eggs laid in genetic clutch	0.62	-1.00	>0.10
Chick-feeding rate of foster father per chick	0.01	-0.06	>0.50
Chick-feeding rate of foster mother per chick	0.15	-1.20	>0.50
Total chick-feeding rate of foster parents per chick	0.87	0.49	>0.10

Final model: $F = 20.93$, d.f. = 2, 14, $r = 0.84$, $P < 0.0001$.

The model is based on 17 observations. The dependent variable was the son's badge size. A forward stepwise procedure was used in which independent variables were included in the final model only if they explained a significant portion of the residual variation in the dependent variable. Independent variables were subsequently removed from the final model if they no longer increased the amount of variation explained in the dependent variable by a significant factor. For independent variables in the final model, F -ratio values show the effect of removing that variable from the model. For independent variables not entered into the final model, F -ratio values show the effect of entering that variable into the model. Coefficient values show the partial regression coefficients. P -values are the associated probabilities.

sons, although this model should be interpreted cautiously because it is based on few observations.

The results of our study are contrary to the assumption of all Fisherian and many good-genes models of sexual selection that sexual ornaments are largely determined by heritable genetic factors¹. Instead, our results show that non-genetic factors can be significant. Our findings are therefore in agreement with the idea that the expression of some sexual ornaments is determined either by environmental factors or, more likely, by an interaction between genetic and environmental factors^{8,16–19}.

Because we have data from only a single population from a single year, it is not possible to predict the generality of this result. We do not suggest that sexually selected traits have zero heritability: the phenotypic variation in most traits is likely to be heritable to some extent³. We do suggest, however, that it is premature to assume that non-genetic effects are rare or unimportant with respect to the expression of sexual ornaments in species showing parental care. Indeed, even in the two species upon which this assumption is based^{5,6}, the pattern now appears more complex than originally assumed. In the great tit, *Parus major*, in which the width of the breast stripe is thought to be sexually selected, a parallel study on another population revealed no significant heritability of breast-stripe size, but found a substantial environmental contribution to variation in this trait⁷. Similarly, in the case of the collared flycatcher, *Ficedula albicollis*, analyses of a 14-year database have shown that the heritability of the sexually selected white forehead patch was dependent on a range of environmental factors⁸. In both these cases, subsequent studies have shown that, in common with many other studies²⁸, heritability varies considerably between years, and often seems to covary with environmental conditions. It would therefore seem prudent to include genotype–environment interaction in future models of sexual selection, particularly when they are likely to be applied to species in which environmental sources of variation may be common, such as those displaying parental care. Indeed, the neglect of gene–environment interactions seems puzzling given the frequency with which empirical studies of sexual selection have demonstrated that the expression of sexual ornaments is influenced by the environment^{1,17}. □

Methods

Assignment of paternity. Putative parents were identified from their unique colour-band combinations during behavioural observations of adults feeding nestlings. The genetic paternity of all offspring was confirmed by genotyping these putative parents alongside the offspring using polymerase chain reaction (PCR) amplification of an array of microsatellite loci²¹, giving a cumulative

probability of false paternal inclusion of 0.008. Across the whole population, fewer than 1% of offspring (3 out of 305) resulted from extra-pair fertilizations. All the offspring in this analysis were assigned to their true genetic fathers.

Badge measurement. Using Vernier calipers, badge size was measured (to the nearest 0.1 mm) from the base of the bill to the point on the breast at which the black feathers ceased to grow. This measure of the 'hidden' badge was robust against the variable size of the visible badge, as buff feathers, which initially conceal the badge, wear away²⁴. Each time a male was captured, badge length was measured three times, from which we calculated a mean measure that was highly repeatable throughout the year within a moult ($r = 0.76$, $F_{12,13} = 7.78$, $P < 0.001$). This linear measurement was used in preference to the area of the badge, which was less repeatable, owing to the inaccuracy of measuring the width across the whole black area at the base of the feathers. Badge length is also the main determinant of overall badge area, explaining more than 70% of badge area ($r = 0.82$, $n = 48$, $P < 0.001$).

Cross-fostering. Whole broods were swapped on the day that the last chick hatched. Broods were swapped randomly with others that were, as far as possible, of the same size and hatching date. Of 87, 48 were swapped with broods of similar size and only 10 out of 87 broods were enlarged or reduced by more than one chick. Blood samples were taken from the brachial vein of chicks 12 days after hatching, and they were ringed with a metal ring (supplied by the British Trust for Ornithology). In October, after the annual moult, offspring were recaptured using mist nets. The badge size of surviving male offspring was estimated using the same procedure as used for adults.

Environmental variables. All nests were monitored throughout the breeding season. Clutch size was the number of eggs incubated by a female. Hatching date was taken as the number of days from 1 May. Tarsus length (to the nearest 0.1 mm) and body mass (to the nearest 0.25 g) were measured 12 days after hatching. Residual mass was estimated as the residual from the regression of body mass on tarsus length. Feeding rates of males and females were scored simultaneously and were mean values from several 20-min observation periods randomly distributed throughout the 16-day nestling period. Behavioural watches were conducted by a random observer using a telescope from an observation point at least 10 m from the nest. All watches were conducted between 4:00 and 8:00 GMT. Brood survival was monitored throughout the rearing period, allowing the calculation of feeding rates per nest per chick.

Received 15 February; accepted 7 June 1999.

- Andersson, M. *Sexual Selection* (Princeton Univ. Press, 1994).
- Pomiankowski, A. & Møller, A. P. A resolution to the lek paradox. *Proc. R. Soc. Lond. B* **260**, 21–29 (1995).
- Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, Massachusetts, 1998).
- Price, T. in *Maternal Effects as Adaptations* (eds Mousseau, T. A. & Fox, C. W.) 202–226 (Oxford University Press, 1998).
- Norris, K. Heritable variation in a plumage indicator of viability in male great tits *Parus major*. *Nature* **362**, 537–539 (1993).
- Sheldon, B. C., Merilä, J., Qvarnström, A., Gustafsson, L. & Ellegren, H. Paternal genetic contribution to offspring condition predicted by size of male secondary sexual character. *Proc. R. Soc. Lond. B* **264**, 297–302 (1997).
- Lemel, J. *Evolutionary and Ecological Perspectives of Status Signalling in the Great Tit (Parus major L.)*. Thesis, Göteborgs Univ., Sweden (1993).
- Qvarnström, A. Genotype-by-environment interactions in the determination of the size of a secondary sexual character in the collared flycatcher (*Ficedula albicollis*). *Evolution* (in the press).
- Boag, P. T. The heritability of external morphology in Darwin's ground finch (*Geospiza*) on Isla Daphne Major, Galapagos. *Evolution* **37**, 877–894 (1983).
- Møller, A. P. Natural and sexual selection on a plumage signal of status and on morphology in house sparrows, *Passer domesticus*. *J. Evol. Biol.* **2**, 125–140 (1989).
- Grant, P. R. The significance of subadult plumage in Darwin's finches *Geospiza fortis*. *Behav. Ecol.* **1**, 161–170 (1990).
- Hill, G. E. Plumage coloration is a sexually selected indicator of male quality. *Nature* **350**, 337–339 (1991).
- Møller, A. P. 1991 Sexual selection in the barn swallow *Hirundo rustica*. I. Determinants of ornament size. *Evolution* **45**, 1823–1836 (1991).
- Lundberg, A. & Alatalo, R. V. *The Pied Flycatcher* (Poyser, London, 1993).
- Williams, J. D., Krueger, J. E. & Harmel, D. H. Heritabilities for antler characteristics and body weight in yearling white-tailed deer. *Heredity* **73**, 78–83 (1994).
- Andersson, M. Evolution of condition-dependent sex ornaments and mating preferences: sexual selection based on viability differences. *Evolution* **40**, 804–816 (1986).
- Rowe, L. & Houle, D. The lek paradox and the capture of genetic variation by condition dependence. *Proc. R. Soc. Lond. B* **263**, 1415–1421 (1996).
- Iwasa, Y. & Pomiankowski, A. Good parent and good gene models of handicap evolution. *J. Theor. Biol.* (in the press).
- Price, T., Schluter, D. & Heckman, N. E. Sexual selection when the female directly benefits. *Biol. J. Linn. Soc.* **48**, 187–211 (1993).
- Wolf, J. B., Moore, A. J. & Brodie, E. D. The evolution of indicator traits for parental quality: the role of maternal and paternal effects. *Am. Nat.* **150**, 639–649 (1997).
- Griffith, S. C., Stewart, I. R. K., Dawson, D. A., Owens, I. P. F. & Burke, T. Contrasting levels of extra-pair paternity in mainland and island populations of the house sparrow: is there an 'island effect'? *Biol. J. Linn. Soc.* (in the press).

22. Veiga, J. P. Badge size, phenotypic quality, and reproductive success in the house sparrow. *Evolution* **47**, 1161–1170 (1993).
23. Møller, A. P. Social control of deception among status signalling house sparrows *Passer domesticus*. *Behav. Ecol. Sociobiol.* **20**, 307–311 (1987).
24. Møller, A. P. Variation in badge size in male house sparrows *Passer domesticus*: evidence for status signalling. *Anim. Behav.* **35**, 1637–1644 (1987).
25. Møller, A. P. Badge size in the house sparrow *Passer domesticus*: effects of intra- and intersexual selection. *Behav. Ecol. Sociobiol.* **22**, 373–378 (1988).
26. Møller, A. P. Frequency of female copulations with multiple males and sexual competition. *Am. Nat.* **139**, 1089–1101 (1992).
27. Veiga, J. P. & Puerta, P. Nutritional constraints determine the expression of a sexual trait in the house sparrow, *Passer domesticus*. *Proc. R. Soc. Lond. B* **263**, 229–234 (1996).
28. Hoffmann, A. A. & Merila, J. Heritable variation and evolution under favourable and unfavourable conditions. *Trends Ecol. Evol.* **14**, 96–101 (1999).

Acknowledgements. We thank NERC (UK) for support and the Landmark Trust for permission to work on Lundy; B. Sheldon for help and advice; the Department of Animal Ecology, Uppsala University for support to S.C.G.; the Department of Zoology, University of Oxford for sabbatical hospitality to I.P.F.O.; D. Dawson, I. Hartley, A. Møller, A. Pomiankowski, D. Richardson, D. Ross and I. Stewart for help and discussion; and R. Alatalo, M. Andersson, P. Harvey, J. Merila, K. Norris, J. Pemberton, A. Pomiankowski, T. Price and A. Qvarnström for comments on the manuscript.

Correspondence and requests for materials should be addressed to S.C.G. (e-mail: simon.griffith@zoologi.uu.se).

Neuronal pacemaker for breathing visualized *in vitro*

Naohiro Koshiya & Jeffrey C. Smith

Cellular and Systems Neurobiology Section, Laboratory of Neural Control,
National Institute of Neurological Disorders and Stroke,
National Institutes of Health, Bethesda, Maryland 20892, USA

Breathing movements in mammals arise from a rhythmic pattern of neural activity, thought to originate in the pre-Bötzinger complex¹ in the lower brainstem. The mechanisms generating the neural rhythm in this region are unknown^{2–5}. The central question is whether the rhythm is generated by a network of bursting pacemaker neurons coupled by excitatory synapses that synchronize pacemaker activity. Here we visualized the activity of inspiratory pacemaker neurons at single-cell and population levels with calcium-sensitive dye. We developed methods to label these neurons retrogradely with the dye in neonatal rodent brainstem slices that retain the rhythmically active respiratory network. We simultaneously used infrared structural imaging to allow patch-clamp recording from the identified neurons. After we pharmacologically blocked glutamatergic synaptic transmission, a subpopulation of inspiratory neurons continued to burst rhythmically but asynchronously. The intrinsic bursting frequency of these pacemaker neurons depended on the baseline membrane potential, providing a cellular mechanism for respiratory frequency control. These results provide evidence that the neuronal kernel for rhythm generation consists of a network of synaptically-coupled pacemaker neurons.

Although it has long been known that the respiratory rhythm is generated in the brainstem, only recently has the critical locus been localized in the pre-Bötzinger complex, a functionally identified subregion of the mammalian ventrolateral medulla oblongata^{1,5,6} (Fig. 1). Neural activity in this region is essential to maintain rhythmic respiratory activity in brainstem neurons and motor nerves *in vivo*^{7,8} and *in vitro*^{1,9}. *In vitro* brainstem slice preparations from neonatal rodents containing the pre-Bötzinger complex spontaneously generate rhythmic inspiratory activity in cranial hypoglossal motor nerves (XII) whose motor neurons are also contained in the slice^{1,9,10}. Studies with these preparations led to the proposal that the kernel for rhythm generation consists of a population of neurons with intrinsic pacemaker properties, with the activity of these neurons being synchronized by excitatory synaptic interconnections^{1,5,6,11}. This pacemaker-network hypothesis^{4,6} has been difficult to test because the pre-Bötzinger complex lies in the heterogeneous reticular formation, making it difficult to identify the pacemaker neurons. To overcome this problem, we developed

techniques to image pacemaker activity.

We found¹² that membrane permeant calcium-sensitive dyes, particularly Calcium Green-1 AM (CaG), will diffuse into axons and retrogradely label neuron cell bodies after microinjection of dye into neonatal rat slice preparations at selected sites through which axons pass. We microinjected CaG near the midline of the slices (Fig. 1a) to label pacemaker neurons, because rhythmic activity in the network on both sides of the medulla is synchronized, requiring functional coupling through bilateral axonal connections crossing the midline. The dye injection resulted in retrograde fluorescent labelling of neurons bilaterally in the reticular formation. A subpopulation of the labelled neurons (Fig. 1b, c) showed increased fluorescence intensity ($\Delta F/F$; Fig. 1d, e) during the inspiratory phase of network activity, monitored by the hypoglossal nerve discharge (f_{XII}). These rhythmically active fluorescent neurons were located in the slice at depths up to the limits of optical resolution ($\sim 80 \mu\text{m}$) and were confined to a region $\sim 300 \mu\text{m}$ in diameter in the ventrolateral medulla ($n = 27/27$ slices; Fig. 1b).

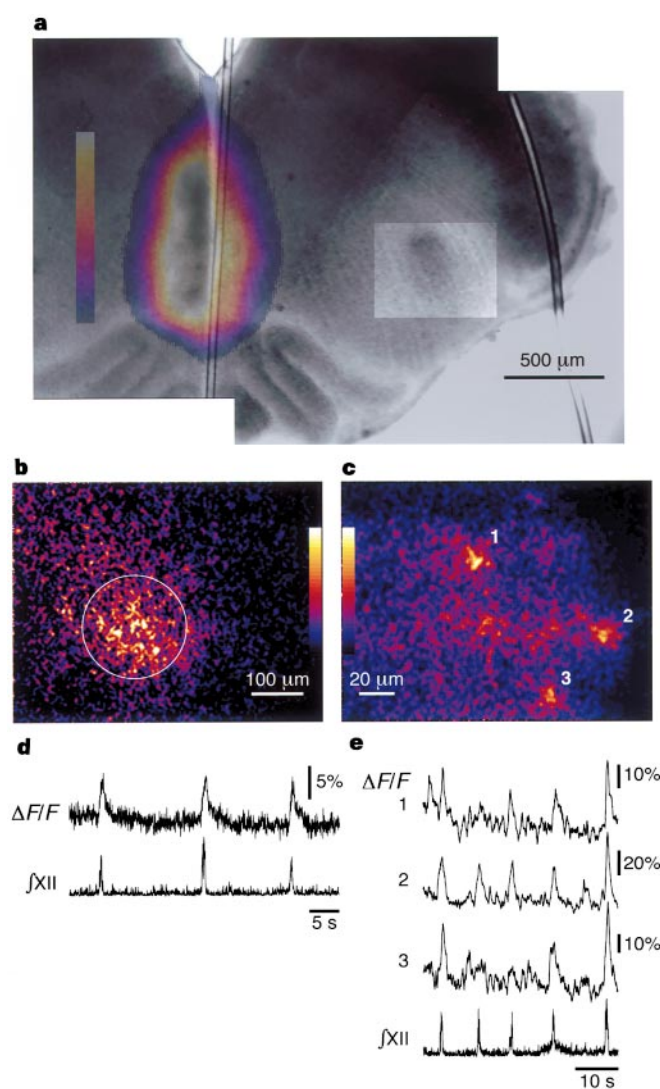


Figure 1 CaG injection site and localized Ca^{2+} activity in pre-Bötzinger complex. **a**, Image of slice with fluorescence (pseudocolour) showing CaG spread from injection site after 12 h, away from Ca^{2+} imaging sites (highlighted rectangle). **b**, Inspiratory 'flash' image (see Methods) in highlighted rectangle in **a**. The elevated fluorescence is localized within the pre-Bötzinger complex¹. **c**, High magnification flash image of region in **b** showing individual inspiratory neurons in a single focal plane (neuron somatas labelled 1–3; see also Fig. 2). **d**, **e**, Rhythmic elevations in $\Delta F/F$ in the circular area in **b**, and (**e**) in individual neurons (1–3; 300 ms moving average) in **c**.

This region corresponded to the area of the pre-Bötzinger complex where respiratory neurons have been mapped electrophysiologically *in vitro*^{1,11,13} and *in vivo*¹⁴.

At the same time, we visualized the morphology of the labelled neurons with infrared-differential interference contrast (IR-DIC) optics, which enabled us to place a patch-clamp pipette on the cell body for electrophysiological recording (Fig. 2A). Under whole-cell current clamp, all neurons examined ($n = 21$ in 15 slices) with rhythmic calcium activities (Fig. 2B) showed rhythmic depolarization and bursts of action potentials during the inspiratory phase (Fig. 2C). The amplitude and timing of the calcium-sensitive fluorescence changes ($\Delta F/F$) varied from cycle to cycle (Fig. 1e), with the onset of the fluorescence transient occurring in some neurons more than 300 ms before the hypoglossal motor discharge (Figs 2–4). The duration of the calcium signal was longer (700–900 ms decay constant) than the period of neuronal spiking and

hypoglossal motor discharge, attributable to slower kinetics of intracellular calcium sequestration after activity-related accumulation^{15,16}. About half of the labelled inspiratory neurons sampled ($n = 11/21$) could burst at frequencies higher than the network frequency when the baseline membrane potential was depolarized under current-clamp conditions, indicating voltage-dependent bursting pacemaker-like properties (Fig. 2C).

To optically identify neurons with pacemaker properties and test the idea that the bursting of pacemaker neurons was synchronized by excitatory synaptic interactions, we imaged the activity of multiple inspiratory neurons simultaneously in single focal planes during blockade of non-NMDA (*N*-methyl-D-aspartic acid) glutamatergic receptors with the antagonist CNQX (6-cyano-7-nitroquinoxaline-2,3-dione; 10–50 μ M). We thought it likely that the synchronizing excitatory interactions would involve non-NMDA receptors, because pharmacological blockade of these receptors, but not NMDA receptors, disrupts the network rhythm⁹. CNQX abolished the inspiratory motor output¹⁹, but around 70% of the labelled inspiratory cells ($n = 23/33$ in three slices) could still generate rhythmic calcium transients. The frequency of the calcium transients depended on the level of neuronal excitation as regulated by the extracellular potassium concentration ($[K^+]_o$, Fig. 3), which is consistent with the voltage-dependent oscillatory bursting behaviour we observed electrophysiologically (Fig. 4). At a given $[K^+]_o$, the neuronal oscillations desynchronized after CNQX (Fig. 5; mean cross-correlation value at zero time lag, 0.735 ± 0.021 s.e. before, and 0.098 ± 0.026 s.e. after CNQX, $n = 14$ pairs of neurons in three slices). This provides direct evidence for a pacemaker network in which excitatory synaptic interactions, mediated by non-NMDA glutamatergic receptors, synchronize cell activity.

We examined the electrophysiological behaviour of single optically identified inspiratory neurons to analyse the pacemaker

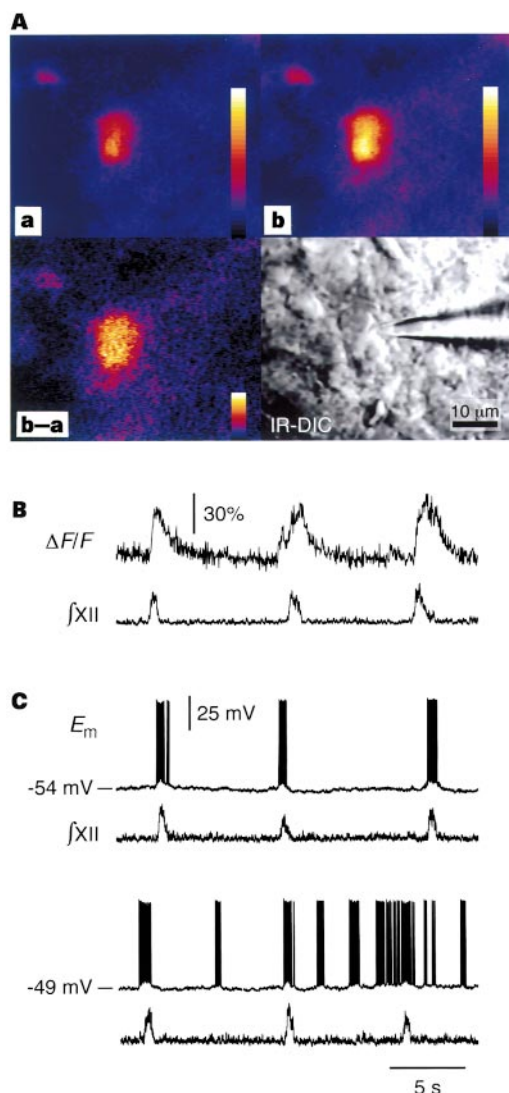


Figure 2 Functional and structural imaging of inspiratory neuron with intrinsic bursting properties. **A**, CaG fluorescence in neurons 60 μ m deep in slice before (**a**) and during (**b**) inspiratory phase. Somata of central neuron in flash image (**b – a**) is shown in IR-DIC structural image, with patch-clamp pipette attached. **B**, Rhythmic elevations in $\Delta F/F$ synchronized with $fXII$. **C**, Whole-cell current-clamp recording of membrane potential (E_m) confirming inspiratory phase depolarization and spiking synchronized with $fXII$. At resting E_m (–54 mV) oscillations were synchronized to network frequency by synaptic inputs. Voltage-dependent intrinsic bursting was revealed by depolarization with constant applied current, causing bursting at a higher frequency.

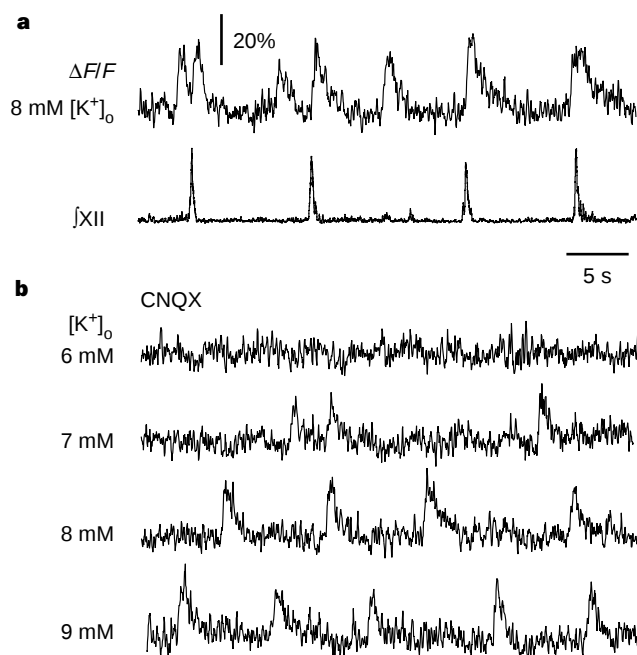


Figure 3 Intrinsic Ca^{2+} activity of inspiratory neuron after blockade of non-NMDA glutamatergic synaptic transmission. **a**, Neuron showing rhythmic $\Delta F/F$ synchronized with $fXII$. At $[K^+]_o = 8$ mM, this neuron showed occasional asynchronous Ca^{2+} transients due to the level of depolarization, indicating intrinsic bursting properties (Fig. 2). **b**, After application of CNQX (10 μ M), neuronal excitability was adjusted with $[K^+]_o$ (6–9 mM). The neuron was quiescent with no CaG fluorescence oscillations at low $[K^+]_o$, and oscillations occurred with increasing frequency as $[K^+]_o$ was elevated. The amplitude of $\Delta F/F$ was reduced after CNQX owing to reduced neuronal burst duration with elimination of synaptic drive (Figs 4, 5).

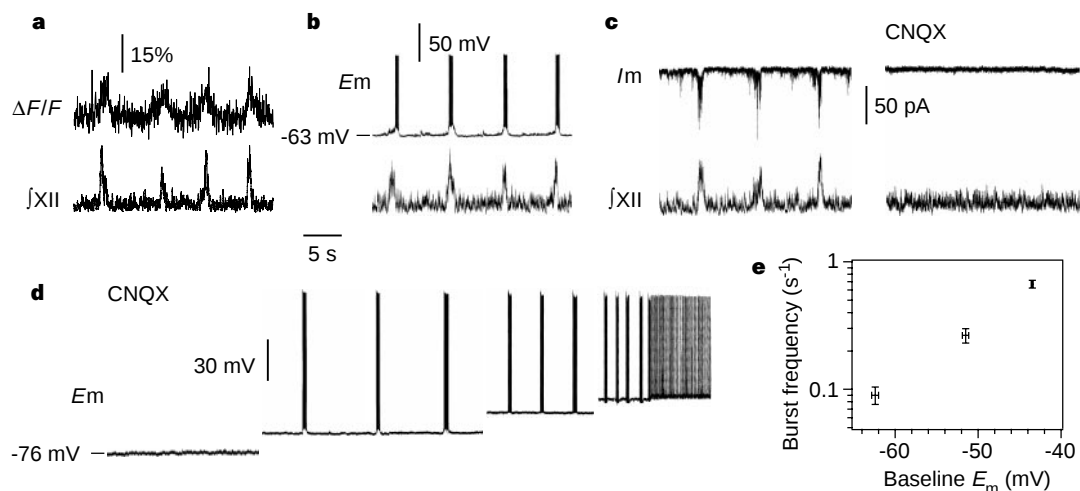


Figure 4 Ca^{2+} activity and voltage-dependent pacemaker properties of inspiratory neuron. **a**, Ca^{2+} activities ($\Delta F/F$) of pacemaker neuron. **b**, Whole-cell current-clamp recording showing rhythmic depolarization of membrane potential (E_m) and bursting synchronous with f_{XII} . **c**, Whole-cell voltage-clamp recording (I_m) at -70 mV holding potential showing inward synaptic currents synchronized with f_{XII} , which were eliminated by CNQX ($10 \mu\text{M}$). **d**, Electrophysiological behaviour

under current clamp after CNQX, showing voltage-dependent oscillatory bursting. **e**, Relationship between burst frequency and baseline membrane potentials (E_m) at three levels of holding current. Error bars indicate standard errors; coefficient of variation of bursting frequency is 0.526, 0.564 and 0.229 (mean: 0.440) at baseline potentials (-62 , -51 and -43 mV, respectively) shown.

properties and synaptic inputs underlying synchronization. All inspiratory cells received phasic excitatory synaptic input during network activity. Pacemaker neurons showed 50–100 pA peak excitatory synaptic currents under voltage clamp at -70 mV (Fig. 4c). These phasic synaptic currents were eliminated by CNQX in all pacemaker neurons examined ($n = 5$ in five slices) as well as in non-pacemaker cells. This confirmed that the desynchronization of rhythmic calcium activities observed with CNQX was due to loss of phasic synaptic inputs. After CNQX treatment, all of the pacemaker neurons exhibited spontaneous, voltage-dependent rhythmic bursting under current clamp where the burst frequency increased monotonically over an order of magnitude (~ 0.05 – 1 s^{-1} ; Fig. 4d, e) as the baseline membrane potential was depolarized over a 20 mV range (~ -65 mV to -45 mV; Fig. 4d, e). The bursting frequency was not completely stationary at a given baseline potential; the mean coefficient of variation (Fig. 4e) ranged

from 0.316 to 0.570 ($n = 5$ neurons). The cells were quiescent at hyperpolarized voltages below -65 mV, and showed tonic spiking ('beating') at the highest depolarization levels (above -45 mV) (Fig. 4d). The other inspiratory neurons examined exhibited only quiescence, or tonic spiking at depolarized levels, after CNQX treatment. These results support the model that the inspiratory neurons with intrinsic oscillatory bursting properties in the pre-Bötzinger complex are a functionally specialized subpopulation involved in generating the network rhythm. We propose that the inspiratory neurons without intrinsic pacemaker properties are 'follower' neurons, transmitting the rhythmic drive to premotor neurons¹⁷ and motor neurons.

The voltage-dependent bursting properties of the pacemaker neurons may be an important cellular mechanism for respiratory frequency regulation; control of baseline membrane potential (for example, by synaptic inputs^{4,6}) can modulate bursting frequency¹⁸.

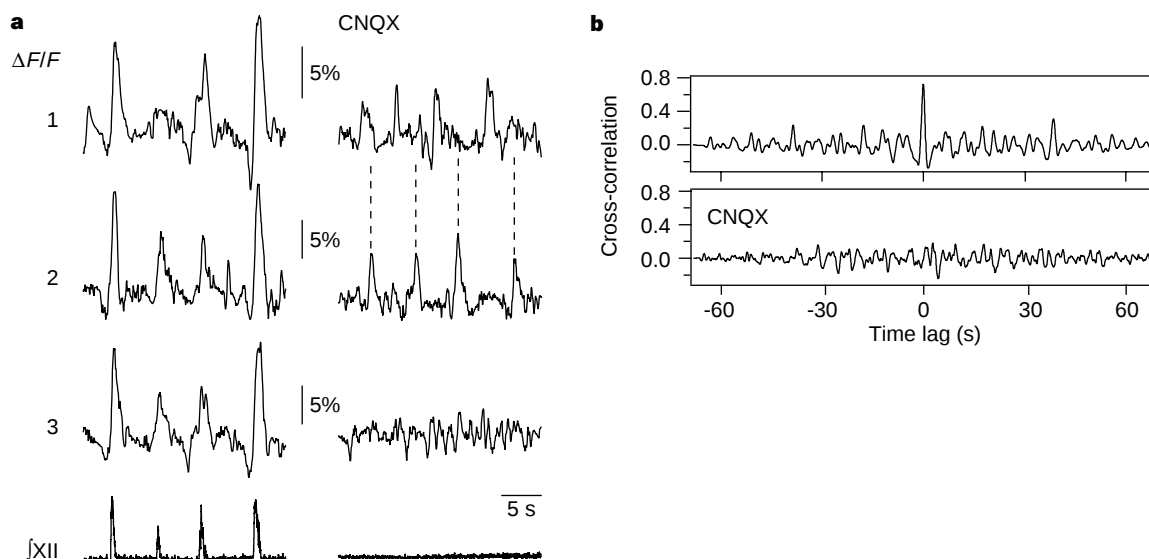


Figure 5 Desynchronization of Ca^{2+} activities of pacemaker neurons after blockade of glutamatergic synaptic transmission. **a**, Synchronized Ca^{2+} activities ($\Delta F/F$) of three inspiratory neurons (1–3) recorded simultaneously in a single focal plane. After CNQX ($50 \mu\text{M}$) two of the cells (1,2) continued to exhibit rhythmic Ca^{2+} transients that were desynchronized (see vertical dashed lines for timing); the

third cell lost Ca^{2+} oscillations. The motor output (f_{XII}) was also eliminated by CNQX. Ca^{2+} signals had noise removed with CWT denoising algorithm. **b**, Cross-correlograms of optical signals from cells 1 and 2 before and after CNQX, showing loss of peak at zero time lag after CNQX, indicating loss of synchrony.

At the network level, the excitatory synaptic interactions, as well as synchronizing cell activity, would act in conjunction with the cellular properties to determine the population oscillation frequency¹⁹. *In vivo*, the pacemaker kernel would be embedded in a more spatially distributed respiratory network that provides additional synaptic inputs for frequency control⁴. Pre-Bötzinger complex neurons have not yet been studied in detail *in vivo*¹⁴ and it will be important to demonstrate the voltage-dependent bursting properties and synaptic interactions in the intact nervous system^{2,20}. Meanwhile, our imaging approach will allow further cellular and network-level analyses of the pacemaker kernel *in vitro*. □

Methods

Slice preparation. Transverse slices (200–400 µm thick) of medulla oblongata were cut from neonatal (P0–P2) rats to contain the pre-Bötzinger complex¹ and rostral end of the hypoglossal motor nucleus with nerve rootlets. The caudal end was cut to expose the pre-Bötzinger complex for imaging. The slice was mounted in a recording chamber (0.2 ml) on the microscope stage and superfused (4 ml min⁻¹) with artificial cerebrospinal fluid¹ (ACSF) including antibiotics (500 units l⁻¹ penicillin, 0.5 mg l⁻¹ streptomycin and 1 mg l⁻¹ neomycin), equilibrated with 95% O₂ and 5% CO₂ (pH = 7.35–7.4 at 27°C). Rhythmic respiratory activity in the hypoglossal nerves (XII) was maintained for up to 48 h by elevating the superfusate K⁺ concentration (7–8 mM)^{1,9}.

Retrograde labelling. Our method, using membrane-permeant Ca²⁺-sensitive dye, was designed to label neurons based on axonal projections and at depth in the slice, without affecting ongoing network activity. Other methods for neuron labelling in tissue with Ca²⁺-sensitive dye, using either bath application of cell-permeant forms²¹ or retrograde transport of impermeant dyes^{21–23}, were not effective. Calcium Green-1 AM (CaG, Molecular Probes; 50 µg) was dissolved in dimethyl sulphoxide (DMSO; 5 µl) containing Pluronic F-127 (BASF; 25 µg) and dispersed in ACSF (10 µl). The CaG solution was micro-injected with a glass pipette (~10 µm tip diameter) at 5–7 contiguous sites near the midline contralateral to the side to be imaged (Fig. 1a). The slice was superfused overnight, allowing transmembrane and intra-axonal dye diffusion for retrograde labelling of cell bodies. We saw dye-labelled axons emanating from the injection site with labelled cells in the medio-lateral reticular formation, including the pre-Bötzinger complex. This intra-axonal diffusion is consistent with other studies where membrane-permeable Ca²⁺-indicator dyes have been used to stain axon terminals anterogradely after local superfusion of slices^{24,25}; microinjection of dye and retrograde labelling of cell bodies, however, have not been previously documented, to our knowledge.

Calcium imaging. We visualized CaG-labelled neurons with a fixed-stage upright videomicroscope system (Axioskop FS, Zeiss) with a 75 W xenon epi-illuminator, optical filters (excitation 485 nm, emission 530 nm and 505 nm beam splitter, Omega Optical) and water-immersion objectives (Zeiss Achromplan). Fluorescence images were captured with a CCD camera fibre-optically coupled to a Gen IV (ref. 26) image intensifier (ICCD-350F/1000F, VideoScope International) and recorded on video tape together with electrophysiological signals. The ICCD allowed low-level epi-illumination (0.3–1%), minimizing phototoxicity and bleaching. Changes in fluorescence intensity ($\Delta F/F$) were detected in real time with an image processor (ARGUS 20, Hamamatsu Photonics) and quantified off-line with NIH Image (<http://rsb.info.nih.gov/ni-h-image/>) after digitizing images using an audio-video board (VideoVision, Radius). Images were formatted as a QuickTime movie file, allowing exact video-frame matching with electrophysiological signals recorded in the audio track. Inspiratory 'flash' images were obtained by subtracting the CaG image at rest from that during inspiration (Fig. 2), and displayed in pseudocolour. We quantified fluorescence signals by averaging the intensity in regions of interest (for example, somata areas) for every video frame. In cases where photon stochastic noise was significant at high image intensification and $\Delta F/F$ was relatively small (4–6%), noise was removed from signals using continuous wavelet transforms (CWT; 0–7 s period- and 50%-pass, 'Paul'; <http://paos.colorado.edu/research/wavelets/>)²⁷, inverse-CWT, and 300 ms moving average. We calculated cross-correlations for simultaneous signals from pairs of neurons for 68.3 s periods (2,048 video frames).

Structural imaging. We obtained videomicroscopic IR-DIC images²⁸ with an extended IR Newvicon camera (Hamamatsu Photonics), and enhanced then

with the ARGUS 20 in real time. Infrared (900 nm) transmission was optimized with a polarizer (wide spectrum, Melles Griot) and analyser (850–950 nm, Polarcor, Corning). For simultaneous imaging, Newvicon and ICCD cameras were mounted on a dual-port imaging head incorporating a dichroic beam splitter²⁹ (700 nm, Zeiss), and images were aligned by reverse scanning and scan range shifting of the Newvicon camera.

Electrophysiology. We recorded voltage- and current-clamp data with an EPC-9 amplifier (version C, HEKA Elektronik) controlled by Pulse software (HEKA; 2.9 kHz low-pass filter). The recording electrodes (3.5–4.5 MΩ) were positioned with computerized microdrives (Märzhauser). The pipette solution was with¹ or without³⁰ calcium chelator. Measured potentials were corrected for the liquid junction potential. Hypoglossal nerve activity was recorded with suction electrodes, amplified and filtered (0.3–2 kHz), then rectified and integrated (fXII) for display; the signals were recorded on VCR tape with patch-clamp recording signals via a PCM digital processor (VR-100B, Instru-tech).

Received 5 October 1998; accepted 25 May 1999.

- Smith, J. C., Ellenberger, H. H., Ballanyi, K., Richter, D. W. & Feldman, J. L. Pre-Bötzinger complex: a brainstem region that may generate respiratory rhythm in mammals. *Science* **254**, 726–729 (1991).
- Richter, D. W., Ballanyi, K. & Schwarzacher, S. Mechanisms of respiratory rhythm generation. *Curr. Opin. Neurobiol.* **2**, 788–793 (1992).
- Ramirez, J. M. & Richter, D. W. The neuronal mechanisms of respiratory rhythm generation. *Curr. Opin. Neurobiol.* **6**, 817–825 (1996).
- Smith, J. C. in *Neurons, Networks, and Motor Behavior* (eds Stein, P., Grillner, S., Selverston, A. & Stuart, D.) 97–104 (MIT Press, Cambridge, Massachusetts, 1997).
- Rekling, J. C. & Feldman, J. L. Pre-Bötzinger complex and pacemaker neurons: hypothesized site and kernel for respiratory rhythm generation. *Annu. Rev. Physiol.* **60**, 385–405 (1998).
- Smith, J. C., Funk, G. D., Johnson, S. M. & Feldman, J. L. in *Ventral Brainstem Mechanisms and Control of Respiration and Blood Pressure* (eds Truett, C. O., Millis, R., Kiwull-Schone, H. & Schlafke, M.) 463–496 (Dekker, New York, 1995).
- Koshiya, N. & Guyenet, P. G. Tonic sympathetic chemoreflex after blockade of respiratory rhythmogenesis in the rat. *J. Physiol. (Lond.)* **491**, 859–869 (1996).
- Ramirez, J. M., Schwarzacher, S. W., Pierrefiche, O., Oliveira, B. M. & Richter, D. W. Selective lesioning of the cat pre-Bötzinger complex *in vivo* eliminates breathing but not gasping. *J. Physiol. (Lond.)* **507**, 895–907 (1998).
- Funk, G. D., Smith, J. C. & Feldman, J. L. Generation and transmission of respiratory oscillations in medullary slices: role of excitatory amino acids. *J. Neurophysiol.* **70**, 1497–1515 (1993).
- Ramirez, J. M., Quellmalz, U. J. A. & Richter, D. W. Postnatal changes in the mammalian respiratory network as revealed by the transverse brainstem slice of mice. *J. Physiol. (Lond.)* **491**, 799–812 (1996).
- Johnson, S. M., Smith, J. C., Funk, G. D. & Feldman, J. L. Pacemaker behavior of respiratory neurons in medullary slices from neonatal rat. *J. Neurophysiol.* **72**, 2598–2608 (1994).
- Koshiya, N. & Smith, J. C. Real-time functional and structural imaging of rhythmically active respiratory neurons with calcium-sensitive dyes and IR-DIC in medullary slices *in vitro*. *Soc. Neurosci. Abstr.* **23**, 1252 (1997).
- Smith, J. C., Greer, J. J., Liu, G. & Feldman, J. L. Neural mechanisms generating respiratory pattern in mammalian brain stem-spinal cord *in vitro*. I. Spatiotemporal patterns of motor and medullary neuron activity. *J. Neurophysiol.* **64**, 1149–1169 (1990).
- Schwarzacher, S. W., Smith, J. C. & Richter, D. W. Pre-Bötzinger complex in the cat. *J. Neurophysiol.* **73**, 1452–1461 (1995).
- Friel, D. D. & Tsien, R. W. Phase-dependent contributions from Ca²⁺ entry and Ca²⁺ release to caffeine-induced [Ca²⁺]_i oscillations in bullfrog sympathetic neurons. *Neuron* **8**, 1109–1125 (1992).
- Li, Y. X., Keizer, J., Stojilkovic, S. S. & Rinzl, J. Ca²⁺ excitability of the ER membrane: an explanation for IP₃-induced Ca²⁺ oscillations. *Am. J. Physiol.* **269**, C1079–C1092 (1995).
- Wilson, C. G., Koshiya, N. & Smith, J. C. Real-time functional imaging of hypoglossal pre-motor and motor neurons in thin medullary slices *in vitro*. *Soc. Neurosci. Abstr.* **23**, 875 (1998).
- Butera, R. J., Rinzl, J. & Smith, J. C. Models of respiratory rhythm generation in the pre-Bötzinger complex: I. Bursting pacemaker neurons. *J. Neurophysiol.* **82**, 382–397 (1999).
- Butera, R. J., Rinzl, J. & Smith, J. C. Models of respiratory rhythm generation in the pre-Bötzinger complex: II. Populations of coupled pacemaker neurons. *J. Neurophysiol.* **82**, 398–415 (1999).
- Bianchi, A. L., Denavit-Saubie, M. & Champagnat, J. Central control of breathing in mammals: neuronal circuitry, membrane properties, and neurotransmitters. *Physiol. Rev.* **75**, 1–45 (1995).
- O'Donovan, M., Ho, S. & Yee, W. Calcium imaging of rhythmic network activity in the developing spinal cord of the chick embryo. *J. Neurosci.* **11**, 6354–6369 (1994).
- McClellan, A. D., McPherson, D. & O'Donovan, M. J. Combined retrograde labeling and calcium imaging in spinal cord and brainstem neurons of the lamprey. *Brain Res.* **663**, 61–68 (1994).
- Fetcho, J. R. & O'Malley, D. M. Visualization of active neural circuitry in the spinal cord of intact zebrafish. *J. Neurophysiol.* **73**, 399–406 (1995).
- Regehr, W. G. & Atluri, P. P. Calcium transients in cerebellar granule cell presynaptic terminals. *Biophys. J.* **68**, 2156–2170 (1995).
- Sabatini, B. L. & Regehr, W. G. Timing of neurotransmission at fast synapses in the mammalian brain. *Nature* **384**, 170–172 (1996).
- Inoue, S. & Spring, K. R. *Video Microscopy: the Fundamentals* 2nd edn (Plenum New York, 1997).
- Torrence, C. & Compo, G. P. A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.* **79**, 61–78 (1998).
- Dodt, H. U. & Ziegler, S. Infrared videomicroscopy: a new look at neuronal structure and function. *Trends Neurosci.* **17**, 453–458 (1994).
- Foskett, J. K. Simultaneous Nomarski and fluorescence imaging during video microscopy of cells. *Am. J. Physiol. (suppl.)* **C566**–C571 (1988).
- Tsubokawa, H. & Ross, W. N. IPSPs modulate spike backpropagation and associated [Ca²⁺]_i changes in the dendrites of hippocampal CA1 pyramidal neurons. *J. Neurophysiol.* **76**, 2896–2906 (1996).

Acknowledgements. We thank M. J. O'Donovan and K. R. Spring for advice on optical recording; W. S. Rasband for NIH Image programs; G. P. Compo for the CWT denoising algorithm; and R. E. Burke and A. Lev-Tov for critical comments on the manuscript.

Correspondence and requests for materials should be addressed to N.K. (e-mail: koshiya@nih.gov).

Temporal dissociation of parallel processing in the human subcortical outputs

Yijun Liu^{*†}, Jia-Hong Gao^{*}, Mario Liotti^{*}, Yonglin Pu^{*} & Peter T. Fox^{*}

^{*} Research Imaging Center, [†] Department of Physiology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78284, USA

Many tasks require rapid and fine-tuned adjustment of motor performance based on incoming sensory information. This process of sensorimotor adaptation engages two parallel subcortical neural circuits, involving the cerebellum and basal ganglia, respectively^{1–10}. How these distributed circuits are functionally coordinated has not been shown in humans. The cerebellum and basal ganglia show very similar convergence of input–output organization^{11,12}, which presents an ideal neuroimaging model for the study of parallel processing at a systems level¹³. Here we used functional magnetic resonance imaging to measure the temporal coherence of brain activity during a tactile discrimination task. We found that, whereas the prefrontal cortex maintained a high level of activation, output activities in the cerebellum and basal ganglia showed different phasic patterns. Moreover, cerebellar activity significantly correlated with the activity of the supplementary motor area but not with that of the primary motor cortex; in contrast, basal ganglia activity was more strongly associated with the activity of the primary motor cortex than with that of the supplementary motor area. These results demonstrate temporally partitioned activity in the cerebellum and basal ganglia, implicating functional independence in the parallel subcortical outputs. This further supports the idea of task-related dynamic reconfiguration of large-scale neural networks^{14,15}.

The lateral cerebellum and basal ganglia appear to have co-operative roles in cognitive processes^{9,10}, probably implemented through hierarchical operations employing parallel information

processing¹⁴. Although neurophysiological and computational models^{3,15} have characterized the temporal properties of neural architecture, the mapping of dynamic processes during behavioural events allows us to model neural systems in an intact human brain, studying both the temporal and spatial dimensions.

We used functional magnetic resonance imaging (fMRI)^{16–18} to examine the time course of brain activation during a tactile object discrimination (TOD) task (see Methods). In the task, a covert judgment was required for matching the shapes of two objects, one in each hand. Meanwhile, perception of the objects relied on sensory inputs that were continuously updated by active finger searching and feeling¹⁹. As the TOD task induced distributed brain activation involving cognitive, somatosensory and motor components, three types of time course were found in the regions showing task-related activity: sustained, slowly rising and slowly falling (Fig. 1). The sustained activation (a flattened time course with short haemodynamic delay¹⁶) was in the dorsolateral prefrontal cortex (PFC). This temporal pattern is consistent with the dynamics of prefrontal activation seen during a working memory task²⁰, and may reflect an executive process for regulating the ‘on-line’ information during the TOD task. At the same time, transient changes in activity were observed in the primary motor cortex (M1) and supplementary motor area (SMA) (agreeing with the time functions reported previously^{20,21}), and in the output nuclei of both the cerebellum and basal ganglia. The time courses of activation showed two distinct patterns in these areas, in which activity in the cerebellar dentate nucleus associated closely with SMA activity, and that of the globus pallidus associated more with M1 activity. These patterns of association are not readily accommodated by existing models of anatomical connectivity in the subcortical outputs^{4–8}.

Tracking time courses of activation in small subcortical nuclei requires good anatomical localization. We used high-resolution fMRI to scan a single brain slice (Fig. 2), a plane through the main axis of the superior cerebellar peduncle and transverse to the output nuclei of both the cerebellum and the basal ganglia. Significant signal changes were found specifically in the internal globus pallidus and in the substantia nigra pars reticulata (Fig. 2a). These two nuclei receive convergent inputs from the striatal spiny neurons and send the basal ganglionic output to the cerebral cortex

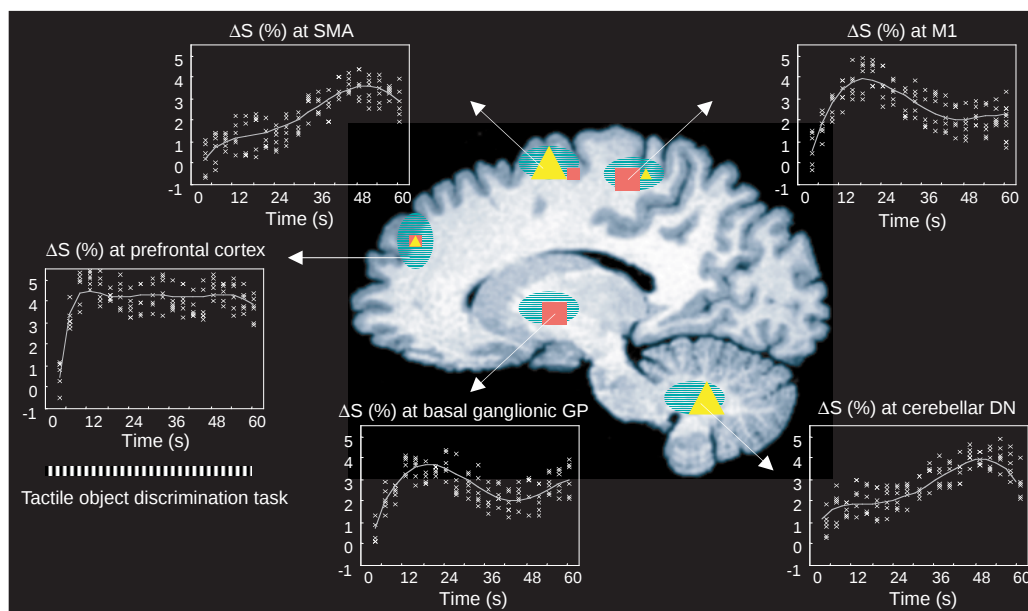


Figure 1 Time courses of EPI-fMRI signal change ($\Delta S = S_{\text{TOD}} - S_{\text{Rest}}$), pooled from six subjects, in the regions showing TOD task-associated activity (see Methods). These regions on the left side of brain were schematically labelled (ovals) by projection to a sagittal image plane ($X = -13$ mm) with reference to the Talairach space. The triangles (squares) mark the regions showing the same

temporal pattern of activation as the dentate nucleus (globus pallidus). The mean ($\pm$ s.e.m., $n = 6$) values of Talairach coordinates (X, Y, Z) for the labelled cortical activation were: PFC ($-38 \pm 2.4, 32 \pm 4.3, 27 \pm 2.7$); M1 ($-35 \pm 2.1, -16 \pm 4.7, 51 \pm 2.3$); SMA ($-3 \pm 2.4, -6 \pm 5.1, 56 \pm 2.2$). DN, dentate nucleus; GP, globus pallidus.

through the thalamus^{6,7,12}. Simultaneously, the cerebellar dentate nucleus and one of its efferent targets, the red nucleus, were also activated. In parallel with the basal ganglia, but with distinct cortical targets, the dentate nucleus relays the lateral cerebellar output from the Purkinje cells in the cerebellar cortex^{8,11}. The active participation of the red and dentate nuclei in tactile discrimination is consistent with a view of sensory modulation of the cerebellar output by a positive feedback loop that includes the cerebellum, red nucleus and inferior olive^{22,23}. The co-activation of these output nuclei, albeit averaged over the entire trial period, indicates that both the cerebellum and basal ganglia are involved in the TOD task.

Based on the activation map (Fig. 2a), the temporal-spatial dependence of activity changes was further revealed by a within-task covariance analysis in which the variation of fMRI signal at each image time was taken into account (see Methods). Although the

output nuclei of both cerebellum and basal ganglia were active during the task, the covariance maps showed a dissociating pattern of this co-activation when the fMRI time series at different regions were used as reference functions (Fig. 2b, c). More specifically, using the within-task time series in the right dentate nucleus as a reference for searching pixel-wise covariance, the left dentate nucleus and both sides of the red nucleus showed significant correlation with the reference, whereas the globus pallidus and substantia nigra did not (Fig. 2b). In contrast, there was no significant correlation in the dentate and red nuclei when the time series in an active region of the globus pallidus was used as a reference (Fig. 2c). Note the covariance between the active regions (defined on Fig. 2a) was driven by the activity within a task block and independent of the block design, so that the autocorrelation imposed by effects from the cyclic experiment paradigm was eliminated²⁴. Therefore, the differential patterns of co-activation in the cerebellum and basal ganglia may represent a temporal parcellation of the activity in these subcortical nuclei.

We analysed echo-planar imaging (EPI) of the whole brain in the same way (Fig. 3). Consistent with the time courses (Fig. 1), the covariance maps showed different patterns of task-related activity in the cortical regions as well as in the subcortical nuclei (Fig. 3b, c). Based on the anatomy of non-human primates, we focused on the areas corresponding to the hand representations in the motor cortex⁵. Apart from some degree of overlap at the thalamus, M1 and the SMA receive segregated inputs from the cerebellum and basal ganglia; M1 receives input mostly from the dentate nucleus, whereas the SMA receives input from the globus pallidus⁶⁻⁸. However, even though both cerebellum and basal ganglia showed activity, the dissociation did not follow the known output pathways to the motor cortex. In particular, the signal change in the SMA, rather than M1, significantly correlated with the reference in the dentate nucleus (Fig. 3b), and M1, not the SMA, was associated with the globus pallidus reference (Fig. 3c).

To demonstrate the temporal-spatial interaction, we performed a covariance analysis across both trials and subjects on a region of interest (ROI) basis (Fig. 4). When the inter-regional correlation was expressed as a function of time during the TOD trial period, the coherence of the fMRI signal between the dentate nucleus and the globus pallidus progressively decreased relative to a resting correlation (Fig. 4a). The loss of the proportional change in fMRI signal in the trials and subjects may also indicate a dissociating effect between the parallel outputs of the active nuclei. The correlation matrices (Fig. 4b) further illustrate the spatial interaction among the four ROIs by averaging the correlation coefficients over trials and subjects. In contrast to the resting condition, the task-related correlation showed a clear pattern of subcortico-cortical interaction that was consistent with the results of the pixel-wise covariance analysis (Fig. 3b). Resting correlations may represent functional connectivity associated with spontaneous fluctuations of activity among anatomically connected regions^{25,26}. In this case, the change in inter-regional covariance may reflect activity-dependent modulation of connectivity as the task progresses. However, the disagreement of the current results with the known anatomy should be interpreted with caution. First, although the globus pallidus has inhibitory outputs, we have found no inverse correlation between it and the SMA (Fig. 4b), indicating that their intrinsic connectivity may not be functionally expressed during the task. In addition, the resting dentate nucleus-globus pallidus correlation probably reflects the activity coherence in the two nuclei rather than a direct connection between them.

Together, our data show a dissociating effect in the regions that are co-activated during the TOD task. The temporal dissociation between the output pathways of the cerebellum and basal ganglia did not follow their anatomical segregation⁶⁻⁹. The changes in inter-regional covariance indicate that the information flow may follow different routing strategies, which may be affected by synaptic efficacy and recurrent signalling during the task^{14,15}. Our results

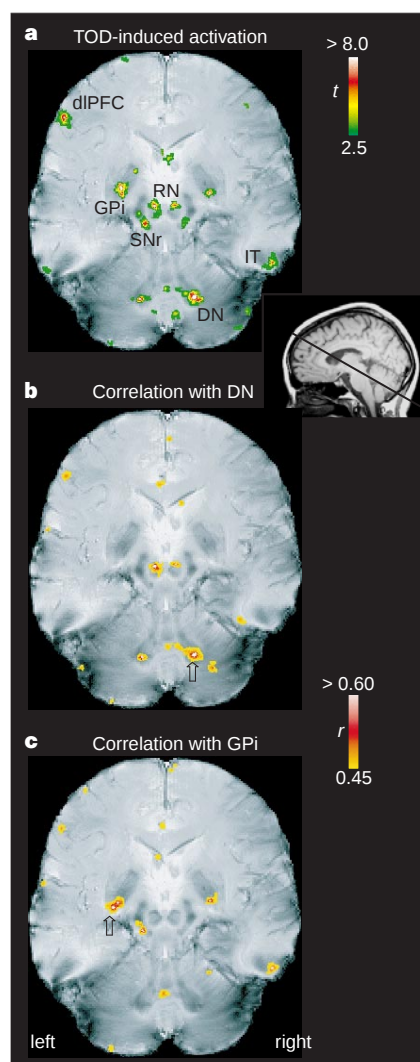


Figure 2 Temporal parcellation of TOD task-associated activity in the output nuclei of cerebellum and basal ganglia. **a**, Activation map overlaid on the T₂-weighted image. **b**, **c**, correlation (*r*) maps showing the time-dependent inter-regional covariance within a task period: **b**, using the time series in an active region of the dentate nucleus (defined in **a**) as a reference function; **c**, using the time series in an active region of the globus pallidus as a reference. The thresholds for the *t*-map and *r*-maps present a statistical significance level ($P < 0.05$) of the mean change in fMRI signal and the correlation (*r* versus zero, $N = 15$) with a pixel-clustering size of six²⁸. dlPFC, dorsolateral PFC; GPI, internal globus pallidus; SNr, pars reticulata of substantia nigra; IT, inferiotemporal cortex. The data are from one of the six subjects scanned by a high-resolution (1 mm²) gradient-echo MRI sequence.

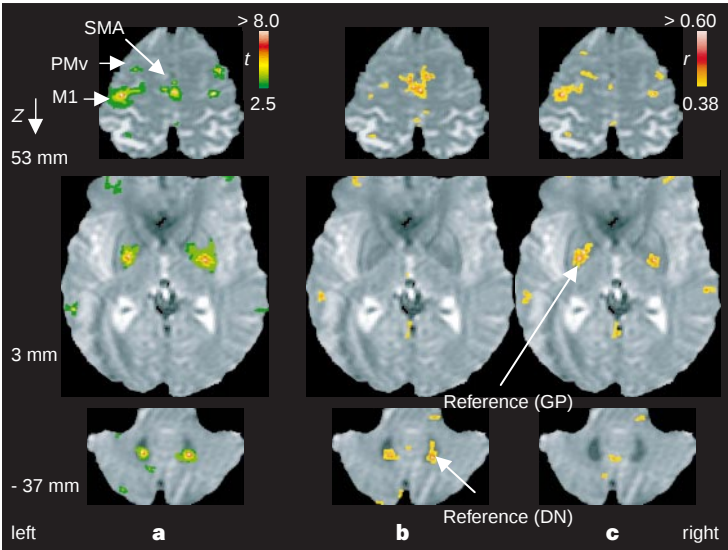


Figure 3 Dynamic reconfiguration of the subcortico-cortical pathways during TOD task. **a**, Activation maps overlaid on the EPI images. **b, c**, Correlation of motor activity with subcortical outputs: **b**, using the time series in the dentate nucleus as a reference; **c**, using the time series in the globus pallidus as a reference. The significance level of the mean change in fMRI signal and the correlation (*r* versus

zero, *N* = 20) was set by the same criteria as in Fig. 2. The Talairach coordinates (*X*, *Y*, *Z*) of the cortical activation (arrows): M1 (−36, −14, 53); PMv (−29, 2, 53); SMA (4, −6, 53). The data are from one of the six subjects in the EPI experiment. PMv, ventral premotor cortex.

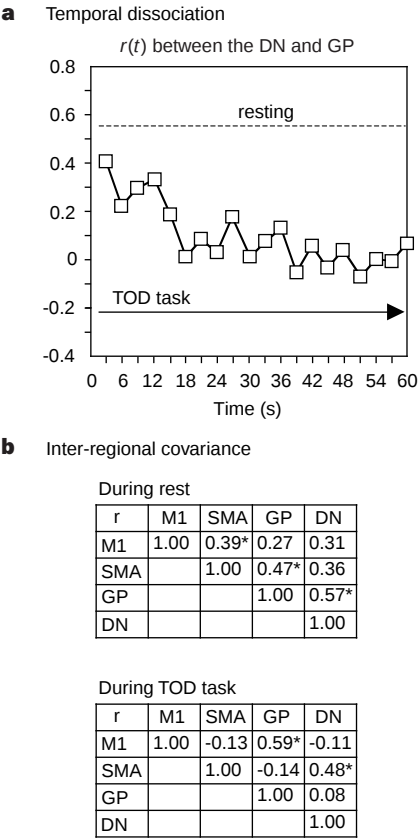


Figure 4 Time-dependent covariance among the motor cortex and the output nuclei of the cerebellum and basal ganglia. **a**, Correlation of EPI-fMRI signal between the dentate nucleus and globus pallidus expressed as a function of time, *r*(*t*). For each image time point within a task period (total 20 EPI images), *r*(*t*) is computed across both trials and subjects (3 × 6). The dashed line indicates a resting level of mean correlation of fMRI signal, computed across an entire period of rest and averaged over the three trials and six subjects. **b**, The mean correlation among four ROIs during rest and TOD. *: *P* < 0.05, a significance level of cross-correlation (*r* versus zero, *N* = 20).

also implicate functional differentiation in the cerebellum and basal ganglia, as temporal independence may allow them to play different roles in the task^{3,10}. The dissociation could come from other potential sources of interference at the level of thalamus or at the input phases of the cerebellum and basal ganglia. However, the cohort of operations in the Purkinje cells or spiny neurons, for example, are very complex^{11,12}, and the current resolutions of fMRI technique do not allow us to address this issue. On the other hand, the temporal coherence of the output activity may be present at a large timescale, reflecting integration in the local circuitry. This timescale is well suited for the imaging study of adaptive sensorimotor processes, such as those underlying tactile discriminations^{22,27}.

Methods

Tasks. The activation paradigm is a time-locked block design consisting of three cycles of rest, control and TOD tasks presented in random order. The tasks were adapted from a previous fMRI study¹⁹. The stimuli were small wooden spheres, differing by zero, one, two or three planar surfaces. In the TOD task, the subject simultaneously grasped two spheres, one with each hand, and used one hand to manipulate the grasped sphere as a reference, while covertly matching the shape of the reference to the sphere grasped by the opposite hand. If the latter stimulus was judged to be different from the reference, it was dropped and another sphere was picked up and compared with the reference. If the two stimuli in both hands were judged to be a match or there was no match after three tests, both the reference and test spheres were released and a new run started with the opposite hand grasping a reference. In the control task, the subject reached and grasped, in pincer fashion, a stimulus at random and then raised and dropped it without active shape discrimination and matching. This sequence was repeated continuously and was performed independently by each hand. At rest, the subject did not touch the stimuli and remained still. The subjects kept their eyes closed during all conditions and task performance was continuously monitored on a video camera. Each subject underwent nine functional scans (see below), corresponding to the nine blocks.

MRI techniques. Experiments were performed on a 2.0 T MRI scanner (Elscent/Prestige). T₂-weighted EPI and conventional gradient-echo fMRI pulse sequences were used. The parameters of the EPI sequence were: TR/TE/θ = 3000 ms/45 ms/90°; slice thickness, 6 mm; in-plane resolution, 2.7 mm × 3.3 mm. For each EPI fMRI experiment, 22 brain slices and 180 images per slice were acquired when the subject performed three cycles of rest, control and TOD. Nine blocks (three blocks for each condition) of data were collected. Each block lasted 60 s and included 20 EPI images. The

parameters of the conventional sequence for the single slice scan were: TR/TE/θ = 60 ms/40 ms/20°; slice thickness, 5 mm; in-plane resolution, 1.0 mm × 1.0 mm. In the conventional gradient-echo fMRI experiment, 135 images were acquired when the subject performed three cycles of rest, control and TOD. Each block lasted 180 s and comprised 15 conventional fMRI images. We acquired anatomical images to localize the subcortical nuclei, which had clear boundaries on the T₂-weighted images.

Data analysis. We used a MEDX software package (Sensor Systems) and in-house programs for image processing, including corrections for head motion and global MRI signal shift. For each voxel, we denote I_0 as the mean MRI signal intensity at the resting condition, which was the average of all images acquired during the three blocks of Rest. The relative MRI signal intensity, $S_{\text{block}}(n)$, during each block was defined as^{17,24} $S_{\text{block}}(n) = [I_{\text{block}}(n) - I_0]/I_0$ ($n = 1, 2, \dots, N$) where $I_{\text{block}}(n)$ is the MRI signal intensity of the n th image in a block (rest, control or TOD), and N is the number of total images acquired in each block. The TOD-induced activation was then determined by comparing the mean of $S_{\text{TOD}}(n)$ and the mean of $S_{\text{Control}}(n)$, using group t -tests and a pixel-clustering technique²⁸. The resulting t -maps and their corresponding T₂ images were then co-registered with the anatomical images using spatial normalization²⁹.

Analyses of the inter-regional covariance were performed on $S_{\text{Rest}}(n)$, and $\Delta S(n) = S_{\text{TOD}}(n) - S_{\text{Rest}}(n)$. At each image in the time series, we calculated $\Delta S(n)$ by paired subtraction between blocks (that is, the first image in TOD minus the first image in Rest, and so on). The time course in an active ROI (defined on the t -maps, see above) was then measured by averaging $\Delta S(n)$ on the number of pixels in the ROI. The resulting spatially averaged time courses of $\Delta S(n)$ in the ROIs (dentate nucleus, globus pallidus, SMA, M1 and PFC) are shown in Fig. 1. The covariance maps were generated on a pixel-by-pixel basis by correlating the time course in a reference ROI (for example, an active region in the dentate nucleus) with those in the pixels throughout the image(s). The cross correlation (r) was calculated using the method from a previous study of fMRI time courses¹⁸. Here we only utilized the time series within the period of a block, and used the time series averaged on an active region as the reference function, a method that is independent of the blocked paradigm. The significance of r (versus zero) was determined by one-tail paired t tests ($N = 20$ or 15). The same spatial normalization procedure and clustering size were used for generating the covariance maps.

Received 5 March; accepted 1 June 1999.

- Allen, G. I. & Tsukahara, N. Cerebrocerebellar communication systems. *Physiol. Rev.* **54**, 957–1006 (1974).
- Marsden, C. D. The mysterious motor function of the basal ganglia: The Robert Wartenberg Lecture. *Neurology* **32**, 514–539 (1984).
- Houk, J. C. & Wise, S. P. Distributed modular architecture linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. *Cereb. Cortex* **5**, 95–110 (1995).
- Ghez, C. in *Principles of Neural Sciences* (eds Kandel, E. R., Schwartz, J. H. & Jessel, T. M.) 609–625 (Elsevier, New York, 1991).
- Rouiller, E. M. in *Hand and Brain* (eds Wing, A. M., Haggard, P. & Flanagan, J. R.) 99–124 (Academic, San Diego, 1996).
- Hoover, J. E. & Strick, P. L. Multiple output channels in the basal ganglia. *Science* **259**, 819–821 (1993).
- Hoover, J. E. & Strick, P. L. The organization of cerebellar and basal ganglia outputs to primary motor cortex as revealed by retrograde transneuronal transport of herpes simplex virus type 1. *J. Neurosci.* **19**, 1446–1463 (1999).
- Middleton, F. A. & Strick, P. L. Dentate output channels: motor and cognitive components. *Prog. Brain Res.* **114**, 553–566 (1997).
- Middleton, F. A. & Strick, P. L. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science* **266**, 458–461 (1994).
- Houk, J. C. On the role of the cerebellum and basal ganglia in cognitive signal processing. *Prog. Brain Res.* **114**, 543–552 (1997).
- Voogd, J. & Ruigrok, T. J. H. Transverse and longitudinal patterns in the mammalian cerebellum. *Prog. Brain Res.* **114**, 21–37 (1997).
- Graybiel, A. M., Aosaki, T., Flaherty, A. W. & Kimura, M. The basal ganglia and adaptive motor control. *Science* **265**, 1826–1831 (1994).
- Churchland, P. S. & Sejnowski, T. J. Perspectives in cognitive neuroscience. *Science* **242**, 741–745 (1988).
- Bressler, S. L. Large-scale cortical networks and cognition. *Brain Res. Rev.* **20**, 288–304 (1995).
- Van Essen, D. C., Anderson, C. H. & Olshausen, B. A. in *Large-Scale Neuronal Theories of the Brain* (eds Kock, C. & Davis, J. L.) 271–299 (MIT Press, Cambridge, Massachusetts, 1994).
- Kwong, K. K. *et al.* Dynamic magnetic resonance of human brain activity during primary sensory stimulation. *Proc. Natl Acad. Sci. USA* **89**, 5675–5679 (1992).
- Ogawa, S. *et al.* Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. *J. Biophys.* **64**, 803–812 (1993).
- Bandettini, P. A., Jesmanowicz, A., Wong, E. C. & Hyde, J. S. Processing strategies for time course data sets in functional MRI of the human brain. *Magn. Reson. Med.* **30**, 161–173 (1993).
- Gao, J. H. *et al.* Cerebellum implicated in sensory acquisition and discrimination rather than motor control. *Science* **272**, 545–547 (1996).
- Cohen, J. D. *et al.* Temporal dynamics of brain activation during a working memory task. *Nature* **386**, 604–608 (1997).
- Courtney, S. M., Ungerleider, L. G., Keil, K. & Haxby, J. V. Transient and sustained activity in a distributed neural system for human working memory. *Nature* **386**, 608–611 (1997).

- Houk, J. C., Keifer, J. & Barto, A. G. Distributed commands in the limb premotor network. *Trends Neurosci.* **16**, 27–33 (1993).
- Liu, Y. *et al.* Involvement of the human red nucleus in sensory discrimination. *Proc. Int. Soc. Magn. Reson. Med.* **6**, 110 (1998).
- Friston, K. J. *et al.* Analysis of fMRI time-series revised. *NeuroImage* **2**, 45–53 (1995).
- Biswal, B., Yetkin, F. Z., Haughton, V. M. & Hyde, J. S. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34**, 537–541 (1995).
- Friston, K. J., Frith, C. D., Liddle, P. F. & Frackowiak, R. S. J. Functional connectivity: the principle-component analysis of large (PET) data sets. *J. Cereb. Blood Flow Metab.* **13**, 5–14 (1993).
- Roland, P. E., Eriksson, L., Widén, L. & Stone-Elander, S. Changes in regional cerebral oxidative metabolism induced by tactile learning and recognition in man. *Eur. J. Neurosci.* **1**, 3–18 (1988).
- Xiong, J., Gao, J. H., Lancaster, J. L. & Fox, P. T. Clustered pixels analysis for functional MRI activation studies of the human brain. *Hum. Brain Mapp.* **3**, 287–301 (1995).
- Fox, P. T. Spatial normalization: origins, applications, and alternatives. *Hum. Brain Mapp.* **4**, 1–2 (1995).

Acknowledgements. We thank J. C. Houk for comments on the manuscript.

Correspondence and requests for materials should be addressed to Y.L. (e-mail: LIUY1@uthscsa.edu).

Sensory experience modifies the short-term dynamics of neocortical synapses

Gerald T. Finnerty*, Langdon S. E. Roberts & Barry W. Connors

Department of Neuroscience, Brown University, Providence, Rhode Island 02912, USA

Many representations of sensory stimuli in the neocortex are arranged as topographic maps. These cortical maps are not fixed, but show experience-dependent plasticity^{1,2}. For instance, sensory deprivation causes the cortical area representing the deprived sensory input to shrink, and neighbouring spared representations to enlarge, in somatosensory³, auditory⁴ or visual cortex⁵. In adolescent and adult animals, changes in cortical maps are most noticeable in the supragranular layers at the junction of deprived and spared cortex^{6–9}. However, the cellular mechanisms of this experience-dependent plasticity are unclear. Long-term potentiation and depression have been implicated^{10–12}, but have not been proven to be necessary or sufficient for cortical map reorganization. Short-term synaptic dynamics have not been considered. We developed a brain slice preparation involving rat whisker barrel cortex *in vitro*. Here we report that sensory deprivation alters short-term synaptic dynamics in both vertical and horizontal excitatory pathways within the supragranular cortex. Moreover, modifications of horizontal pathways amplify changes in the vertical inputs. Our findings help to explain the functional cortical reorganization that follows persistent changes of sensory experience.

Rats obtain detailed sensory information from their facial whiskers¹³. The five rows of whiskers and corresponding whisker barrel rows in the contralateral primary somatosensory cortex are conventionally labelled A to E¹⁴ (Fig. 1a). We altered the sensory experience of rats by trimming bilaterally either the dorsal three rows (A–C) or the ventral two rows (D, E) of whiskers flush with the skin every day for 10 days beginning at P11–15. Thus, the junction between cortex which had lost its sensory input ('deprived') and that which had retained its input ('spared') was always between the C and D barrel rows. A third sham-trimmed group served as a control. We cut brain slices across the rat whisker barrel rows (Fig. 1a, b). Transilluminating slices allowed reliable visual identification of the barrel rows and, thus, of the junction between deprived and spared cortex. Layer 2/3 pyramidal neurons in the deprived, spared and control cortex did not differ in resting membrane potential, input resistance, membrane capacitance and spike threshold (Table 1), indicating that sensory deprivation did

* Present address: Department of Physiology, School of Medical Sciences, University Walk, Bristol BS8 1TD, UK.

not alter intrinsic cellular properties. We studied the short-term dynamics of excitatory synapses onto layer 2/3 pyramidal cells. Vertical or horizontal intracortical pathways were stimulated with three trains of eleven stimuli at 2–40 Hz with 30 s separation between trains (Fig. 1c, d). EPSP amplitudes were normalized with respect to the first response and the last three responses were averaged to estimate the normalized steady-state amplitude (Fig. 2a). EPSP amplitudes during a train depressed at all stimulation frequencies, as found in other cortical areas^{15,16}. We fitted single exponential curves to find the time constant of EPSP amplitude decay (Fig. 2a). The amount of depression increased with stimulation frequency in both horizontal (Fig. 2b, f) and vertical pathways (Fig. 3b).

Sensory deprivation caused complex changes in synaptic dynamics. We compared horizontal inputs from spared and deprived cortex onto deprived layer 2/3 neurons. The amount of steady-state depression was greater in spared-to-deprived inputs compared with deprived-to-deprived inputs at all stimulation frequencies tested ($P = 0.01$, paired t -test, $n = 12$) (Fig. 2b, c). Depression of control cortex was intermediate between that recorded from deprived and spared cortex. We found that deprived-to-deprived inputs depressed significantly less than controls (mean difference, 0.07 ± 0.02 , $P = 0.02$, paired t -test) whereas spared-to-deprived inputs tended to depress more. This indicated that horizontal inputs to deprived cortex were modified by sensory deprivation, but in opposite ways depending on their source. The increase in depression found in the spared-to-deprived input compared to the deprived-to-deprived input was accompanied by a shortening of the time constant of EPSP amplitude decay (mean decrease, $28 \pm 5\%$, $P = 0.03$, paired t -test) (Fig. 2d). A similar analysis of pathways to spared cortex showed that inputs from

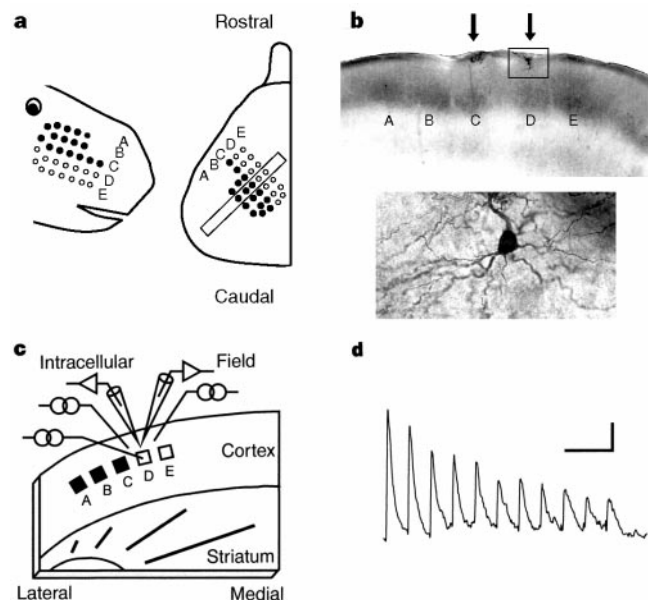


Figure 1 'Across whisker barrel row' brain slice preparation. **a**, Position of whiskers on the rat's snout and their map on the contralateral primary somatosensory cortex. Trimmed whiskers (here, D and E rows) are represented as empty circles. The schematic diagram of the left hemisphere shows the plane of section, 45° to the midline, for preparing brain slices. **b**, Combined biocytin and cytochrome oxidase stain of a slice showing five whisker barrels. Arrows, cell bodies of filled pyramidal cells in layer 2 above C and D barrels. The box enclosing the neuron in the D barrel column is enlarged below. **c**, Arrangement of the recording and stimulating electrodes for the horizontal and vertical pathways on the slice. The barrels are coded as in **a**. The recording protocol is repeated in the same slice after the recording and stimulating electrodes are re-centred above the C barrel column. **d**, Averaged response (three trials) evoked by stimulation at 10 Hz of the vertical layer 4–2 pathway in the C barrel column of the brain slice shown in **b**. Scale bars: 200 ms, 1 mV.

Table 1 Intrinsic properties of layer 2/3 pyramidal cells

	Control	Deprived	Spared
Resting membrane potential (mV)	-79 ± 2	-78 ± 2	-76 ± 1
Input resistance (M Ω)	61 ± 4	53 ± 3	48 ± 5
Membrane capacitance (pF)	260 ± 20	290 ± 20	260 ± 20
Threshold – resting membrane potential (mV)	26 ± 1	25 ± 2	26 ± 2

All data are shown as mean $\pm$ s.e.m. One-way analysis of variance showed no difference in resting membrane potential ($n = 32$, $F_{2,29} = 1.21$, $P = 0.313$), input resistance ($n = 35$, $F_{2,32} = 2.218$, $P = 0.125$), membrane capacitance ($n = 35$, $F_{2,32} = 0.894$, $P = 0.419$) or threshold minus resting membrane potential ($n = 15$, $F_{2,12} = 0.138$, $P = 0.872$).

deprived and spared cortex depressed by comparable amounts and at similar rates (Fig. 2e). The depression of spared inputs was greater than control (mean difference, 0.06 ± 0.02 , $P = 0.04$, paired t -test), and deprived inputs were similar to spared inputs (Fig. 2f).

Sensory deprivation also influenced the dynamics of vertical pathways onto layer 2/3 pyramidal cells. The relationship between normalized steady-state amplitude and frequency was significantly different when comparing the vertical layer 4–2 pathways in deprived and spared cortex ($P < 0.01$, paired t -test, $n = 13$). More detailed analysis revealed that, unlike horizontal inputs to deprived cortex, responses were not modified equally at all stimulation frequencies. Spared cortex depressed significantly more than the deprived cortex at 5 and 10 Hz ($P = 0.043$ and 0.047 , respectively, paired t -tests, $n = 13$) (Fig. 3a), but less difference was seen at 20

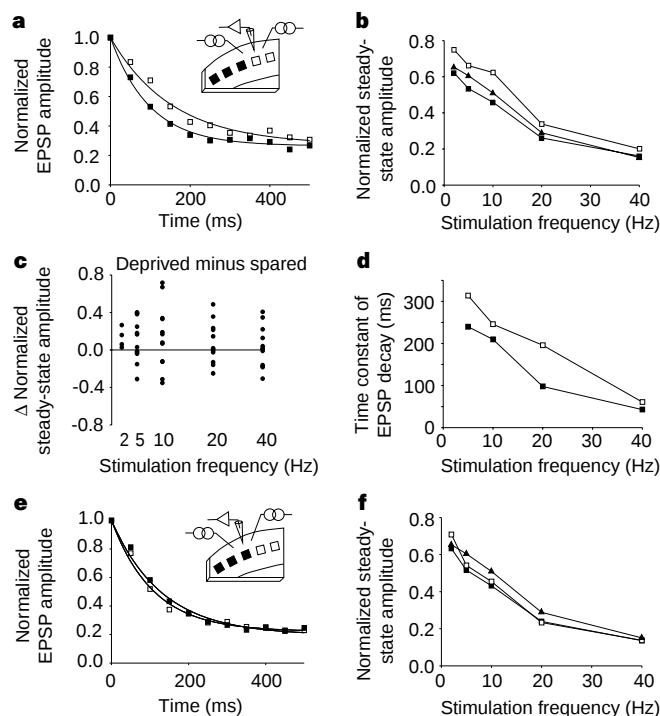


Figure 2 Short-term synaptic dynamics of horizontal pathways. **a**, Normalized EPSP amplitudes in deprived cortex evoked by 20 Hz stimulation of deprived (open squares, $n = 12$) and spared (filled squares, $n = 12$) cortex. Time constants of EPSP amplitude decay were derived from single exponential fits represented by the continuous lines. **b**, Normalized steady-state amplitude plotted against stimulation frequency for inputs to deprived cortex from deprived (open squares, $n = 12$) and spared (filled squares, $n = 12$) cortex compared with a control cortex (filled triangles, $n = 9$). **c**, Differences in normalized steady-state of deprived and spared inputs to deprived cortex from trimmed rats. **d**, Time constant of EPSP amplitude decay for deprived and spared inputs to deprived cortex plotted against frequency. **e**, Normalized EPSP amplitudes in spared cortex evoked by 20 Hz stimulation of deprived (open squares, $n = 12$) and spared (filled squares, $n = 12$) cortex. **f**, Normalized steady-state amplitude plotted against stimulation frequency for inputs to spared cortex from deprived (open squares, $n = 12$) and spared (filled squares, $n = 12$) cortex compared with control cortex (filled triangles, $n = 9$).

and 40 Hz (Fig 3b, c). The responses of control animals were similar to those in deprived cortex, but statistically different from those in spared cortex (mean difference, 0.06 ± 0.02 , $P = 0.04$, paired t -test), indicating that sensory deprivation primarily induced changes in spared cortex. Increased depression in spared inputs compared with deprived inputs was accompanied by a shortening of the time constant of EPSP amplitude decay at 5 and 10 Hz, but not 20 and 40 Hz (Fig. 3d). In four cells (two spared, one deprived and one control) the stimulus train protocol in the layer 4–2 and layer 2–2 pathways was repeated with 200 μ M 2-hydroxysaclofen, a GABA (γ -aminobutyric acid)_B antagonist, in the artificial cerebrospinal fluid (ACSF). The normalized steady-state amplitude did not change significantly in either pathway, indicating that presynaptic GABA_B receptors did not have an important role in stimulus-train-induced depression or the effects of sensory deprivation.

Our data indicated that sensory deprivation modified the synaptic dynamics of both horizontal and vertical pathways in supragranular cortex. Two questions arose. First, could the change in synaptic dynamics contribute to cortical map reorganization in supragranular cortex? Second, what caused the synaptic dynamics to change? To answer these questions we modified and extended a phenomenological model¹⁷. We refer to the EPSP amplitude during a stimulus train as the synaptic efficacy. The data we present have been normalized. In the model, normalized synaptic efficacy depends on stimulation frequency and two other parameters. S is the proportion of available synaptic resources used to generate an EPSP. This proportion can be thought of as the synaptic strength and, for simplicity, that is how we will refer to it. τ is the time constant with which synaptic resources recover after generating an EPSP. Note that we make a clear distinction between synaptic strength, which is a constant during the stimulus train, and synaptic efficacy, which varies. The model predicts that normalized EPSP amplitudes decay exponentially to a steady state during a stimulus train. Therefore, we used the normalized steady-state and time constant of EPSP amplitude decay extracted from our single exponential curve fits to calculate the values for synaptic strength and the recovery time constant at different stimulus frequencies.

We compared horizontal intracortical inputs from spared and deprived cortex onto layer 2 neurons in deprived cortex. The mean synaptic strengths of spared-to-deprived inputs were significantly

greater than those of deprived-to-deprived inputs at all frequencies studied (input from spared cortex, 0.35 ± 0.03 ; deprived cortex, 0.21 ± 0.03 ; $P < 0.001$, paired t -test) (Fig. 4a), with synaptic strengths in control cortex having intermediate values (0.27 ± 0.04). The calculated recovery time constants for deprived, spared and control cortex were not different (input from spared cortex, 480 ± 40 ms; deprived cortex, 520 ± 70 ms; control cortex, 650 ± 100 ms; one-way analysis of variance, $P = 0.25$). We performed a similar analysis of the horizontal intracortical inputs onto layer 2 pyramidal cells in spared cortex. The mean synaptic strengths were similar (spared-to-spared, 0.33 ± 0.07 ; deprived-to-spared, 0.32 ± 0.05 ; $P = 0.67$, paired t -test) with both tending to be bigger than controls (inputs from spared cortex, $P = 0.06$; deprived cortex, $P = 0.05$) (Fig. 4a). The mean recovery time constants were not statistically different (spared cortex, 590 ± 130 ms; deprived cortex, 510 ± 60 ms; $P = 0.21$, paired t -test). We made an independent measurement of the recovery time constant by presenting test responses after the stimulus train. The mean recovery time constant was similar to that extracted from stimulus train data (mean $\tau = 660 \pm 110$ ms, $n = 5$).

In the vertical layer 4–2 pathway, there was a significant increase in synaptic strength in spared cortex, but not in deprived cortex, when compared to controls (spared, 0.34 ± 0.07 vs control, 0.28 ± 0.06 ($P = 0.01$, paired t -test); deprived, 0.29 ± 0.06). The differences in the normalized steady-state of vertical pathways in deprived and spared cortex showed frequency selectivity. Therefore, we compared the synaptic strengths of vertical pathways derived from single stimulation frequencies. We found a statistically significant difference in synaptic strength at 10 Hz (spared, 0.30 ± 0.04 ; deprived, 0.23 ± 0.02 ; $P = 0.03$, paired t -test). The mean recovery time constant did not change with sensory deprivation (spared, $1,020 \pm 210$ ms vs. deprived, $1,190 \pm 180$ ms; $P = 0.554$, paired t -test).

Our results indicate that an increase in synaptic strength accounts for the changes in synaptic dynamics that we found experimentally. We have no data on how synaptic strength changes over the trimming period. However, 2-deoxyglucose studies of trimmed rats indicate increased activity in spared cortex within hours of

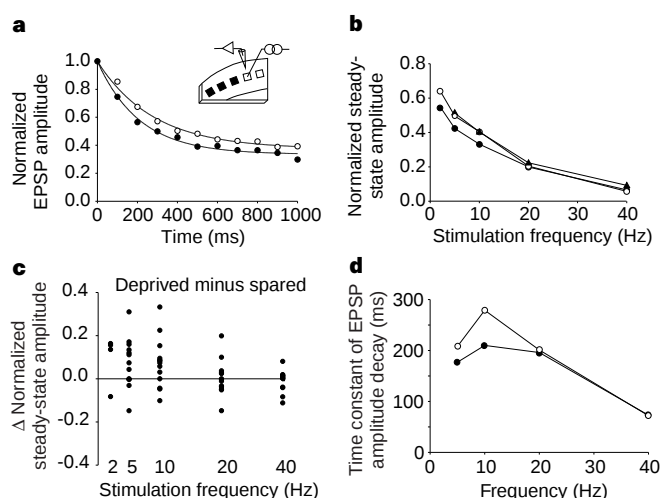


Figure 3 Short-term synaptic dynamics of vertical pathways. **a**, Normalized EPSP amplitudes evoked by 10 Hz stimulation in deprived (open circles, $n = 12$) and spared (filled circles, $n = 12$) cortex. **b**, Normalized steady-state amplitude plotted against frequency for deprived (open circles, $n = 13$), spared (filled circles, $n = 13$) and control (filled triangles, $n = 9$) inputs. **c**, Difference between normalized steady-states of deprived and spared cortex for each pair of cells from trimmed rats. **d**, Time constant of EPSP amplitude decay for layer 4–2 pathways in deprived and spared cortex.

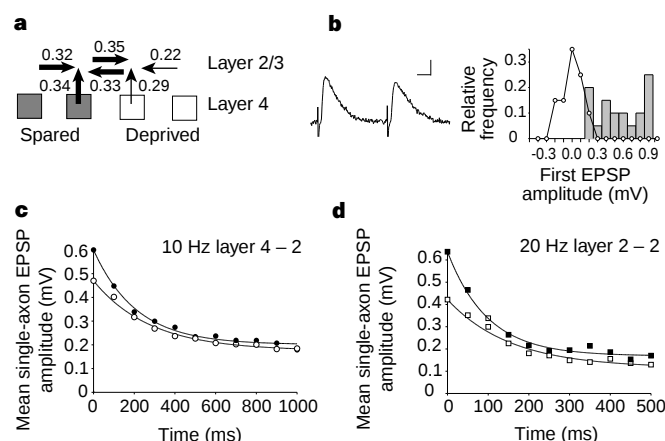


Figure 4 Sensory deprivation changes synaptic strength and synaptic efficacy. **a**, Synopsis of mean synaptic strengths over 5–40 Hz for vertical and horizontal pathways around the junction of spared and deprived cortex. Tail of the arrow, site of stimulation; tip, recording site. Width of arrows scaled with respect to control values for the horizontal (0.27) and vertical (0.28) pathways. **b**, Example of the mean response of 20 successive trials using minimal stimulation protocol on the layer 4–2 pathway. The amplitudes of the first EPSP (grey bars) and noise (open circles) are shown. Scale bars: 20 ms, 0.2 mV. **c**, Estimated synaptic efficacy for an average single-axon response to a 10-Hz stimulus train in the layer 4–2 pathways of deprived (open circles) and spared (filled circles). **d**, Estimated single-axon synaptic efficacy of the deprived-to-deprived (open circles) and spared-to-deprived (filled circles) horizontal pathways in response to a 20-Hz train.

trimming¹⁸ and, in adults, single-unit recordings show increased firing of layer 4 cells by the third day post-trim¹⁹. Thus, we believe that an increase in synaptic strength is consistent with sensory deprivation strengthening that pathway rather than preventing it from becoming weaker as the rat ages. The opposite, sensory deprivation weakening inputs, probably applies to horizontal pathways in deprived cortex where there is a decrease in cell firing at points away from the junction between deprived and spared cortex⁷. The pattern of our results for the horizontal pathways onto deprived cortex and the vertical pathway in spared cortex (increased synaptic strength and unchanged recovery time constant) is the same as that reported for long-term potentiation (LTP) induced by pairing presynaptic firing with postsynaptic depolarization *in vitro*¹⁵. This raises the possibility that LTP may have caused the change in synaptic dynamics that we found. However, this remains to be determined. The horizontal pathway from spared to deprived cortex, which shows the greatest change in synaptic strength, is the pathway where pronounced anatomical changes are found after prolonged sensory deprivation^{20,21}. This indicates that the same mechanisms that modify short-term synaptic dynamics may also drive the longer-term anatomical changes.

Understanding how a change in short-term synaptic dynamics could lead to functional reorganization of cortical maps requires a comparison of the synaptic efficacies of different pathways when the same number of axons have been stimulated. We reasoned that if we found the mean single-axon EPSP amplitude then we could use our stimulus train data to estimate the single-axon synaptic efficacy for each pathway. Therefore, after completing our stimulus train protocol in a pathway, we measured a single-axon evoked EPSP in the same pathway using a minimal stimulation paradigm. This was done for eight pairs of cells (five trimmed and three control slices) in both vertical and horizontal pathways. There are technical and interpretational problems associated with minimal stimulation. However, we were still able to draw some simple conclusions using these data.

Single-axon EPSPs varied greatly between cells, as expected²² (Fig. 4b). However, there was a strong correlation between the computed synaptic strength of the pathways used for minimal stimulation and their measured mean single-axon EPSP amplitudes ($n = 8$, synaptic strengths at 20 Hz, $r = 0.94$, $P < 0.01$, Pearson product moment correlation). We used these data to estimate the synaptic efficacy of mean single-axon inputs onto layer 2/3 cells. The results showed that pathways with larger synaptic strengths were changed in two ways: the amplitude of the initial EPSP in the train was increased, and the EPSP amplitude depressed faster during the stimulus train (Fig. 4c, d). The net effect, comparing vertical layer 4–2 pathways, was that EPSPs in spared cortex were larger than those in deprived and control cortex throughout a 10 Hz stimulus train (Fig. 4c). When we compared horizontal deprived-to-deprived and spared-to-deprived pathways, we found that layer 2–2 EPSPs from spared cortex remained larger throughout a 20 Hz stimulus train (Fig. 4d). At higher frequencies, EPSP amplitudes cross over during the stimulus train, owing to the greater depression of the pathway with higher synaptic strength. However, steady-state EPSPs approach the same values as noted with paired recordings¹⁵.

Our results show, for the first time, to our knowledge, that sensory experience modifies the short-term dynamics of excitatory synapses in the neocortex. Furthermore, our findings might help to explain how modifications of excitatory circuitry in supragranular cortex may contribute to expansion of cortical maps after sensory deprivation lasting days. Recordings in primary somatosensory cortex from freely moving rats using their whiskers reveal a 7–12 Hz oscillation synchronized with whisker protraction²³. Typically, whisker deflection causes neurons in layer 4 to fire 0–2 spikes. Thus, sensory deprivation increases the synaptic efficacy of the layer 4–layer 2 pyramidal cell pathway in spared cortex during whisking and, therefore, the likelihood that spared layer 2 neurons will fire. The horizontal intracortical pathways amplify the changes in the

vertical pathway inputs for two reasons. The synaptic efficacy of the horizontal pathway from spared cortex to deprived cortex is increased, and the synaptic efficacy of inputs to layer 2 neurons in deprived cortex from other pyramidal cells in deprived cortex is decreased. The result is that layer 2 neurons in deprived cortex receive a much larger proportion of their synaptic input from layer 2 neurons in spared cortex, compared with control conditions. Thus, firing of layer 2 neurons in spared and deprived cortex is more likely to be synchronized and the cortical representation of spared inputs expands. The corollary of our sensory deprivation protocol is a perceptual learning paradigm where practice on a task leads to improved discrimination. Cortical maps also reorganize during perceptual learning². Mechanisms similar to those that we have described here may be involved. We suggest that our findings may exemplify a more general mechanism that regulates the allocation of cortical space in primary sensory cortex. □

Methods

Slice preparation. On the day of recording, rats had all whiskers trimmed acutely to blind the electrophysiologist to the rat's sensory history. Brain slices 400 μm thick were cut at 45° to the midline (Fig. 1), stored in an incubation chamber at 34 °C for 1 h and then checked for the presence of five whisker barrel rows¹⁴. Suitable slices were transferred to a submersion chamber at room temperature and bathed in ACSF flowing at 2 ml min⁻¹ and containing (in mM): NaCl 124, KCl 3, NaH₂PO₄ 1.25, NaHCO₃ 26, dextrose 10, MgCl₂ 1 and CaCl₂ 2, and saturated with 95%/5% O₂/CO₂. Ultrasmall bipolar stimulation electrodes (25 μm tip, FHC) were located in layer 4 and layer 2/3 (Fig. 1c). Field potentials in layer 2/3 were recorded with micropipettes (0.5–1 M Ω) filled with 2 mM bicuculline methiodide dissolved in ACSF. We recorded EPSPs from neurons in the immediate vicinity of the field potential electrode with sharp microelectrodes (70–100 M Ω) filled with 3 M potassium acetate and 1% biocytin. An EPSP of ~5 mV was evoked to prevent activation of voltage-gated dendritic channels that could modify EPSP amplitude²⁴. We ensured that fast IPSPs were blocked by stimulating either layer 2 or layer 4 while depolarizing layer 2/3 neurons with current pulses and checking for the presence of a hyperpolarizing potential overlapping the EPSP²⁵. NMDA (*N*-methyl-D-aspartate) receptors were blocked with 50 μM D,L-2-amino-5-phosphopentanoic acid (AP5, Sigma) in the ACSF. We estimated the recovery time constant for short interstimulus intervals by fitting a single exponential to test responses ≤ 500 ms from the end of the stimulus train. When recordings from the first barrel column were completed, we moved stimulating and recording electrodes to the other barrel column at the junction of deprived and spared cortex and repeated the recording protocol. The slice was fixed after recording was completed and the cell type and location of recorded neurons were confirmed histologically using a combined biocytin and cytochrome oxidase stain (Fig. 1b). Synaptic responses were digitized at 10 kHz. Data acquisition and signal analysis were done with Labview (National Instruments). Single exponential curve fitting was done using the Marquardt–Levenberg algorithm in SigmaPlot (SPSS).

Minimal stimulation. Current stimulus intensity was reduced to evoke a single-axon response^{26–28}. Twenty trials of paired stimuli (100 ms interstimulus interval) given every 10 s were averaged to ensure that the mean first and second EPSPs had the same latency and time to peak. Measurements averaged over a 2 ms window were then made on successive traces of EPSP amplitude and baseline noise. We corrected for error in the mean amplitude due to baseline variability by subtracting the baseline mean from the amplitude mean. Single-axon EPSP amplitudes from each pathway were averaged to calculate the mean single-axon EPSP amplitude. The mean synaptic strength of the smaller number of pathways used for minimal stimulation was different on occasions from the mean synaptic strength calculated from pooled stimulus train data. To correct for this, we multiplied mean single-axon EPSPs by the ratio of the synaptic strength calculated from pooled stimulus train data for that pathway to the mean synaptic strength of the pathways used for minimal stimulation.

Model of synaptic depression. The model¹⁷ proposes that a pathway possesses a fixed amount of synaptic resources that exist in one of three states: recovered, active and inactive. An EPSP is modelled by the transfer of a constant proportion of synaptic resources S from the recovered pool into

the active pool. These resources then relax into the recovered pool through the inactive pool. We assume that the decay from active to inactive pools is rapid so that the effective time constant for relaxation into the recovered pool is τ . For simplicity, we express the total amount of synaptic resources as the maximal EPSP amplitude A , which is the magnitude of the evoked EPSP when all the synaptic resources are used. EPSP amplitudes during a stimulus train are given by the product of the synaptic strength and the synaptic resources in the recovered pool at the time of the action potential. Combining the model with the experimental observations that there is a linear relationship between the dendritic excitatory postsynaptic current (EPSC) and somatic EPSP when the somatic EPSP is less than ~ 5 mV²⁴, and that somatic EPSPs add linearly²⁹, we derived an equation relating normalized EPSP amplitudes during a stimulus train. After k stimulus intervals of a regular stimulus train of frequency f , normalized EPSP _{k} = $\{S \cdot e^{-(1/\tau - f \cdot \ln(1-S))t} + (e^{1/(f\tau)} - 1)/(e^{1/(f\tau)} - (1-S))\}$. Thus, a continuous curve drawn through normalized EPSP amplitudes evoked by the stimulus train has an exponential form with a steady state of $(e^{1/(f\tau)} - 1)/(e^{1/(f\tau)} - (1-S))$ and a time constant of decay of $(1/\tau - f \cdot \ln(1-S))^{-1}$.

Received 22 March; accepted 3 June 1999.

- Buonomano, D. V. & Merzenich, M. M. Cortical plasticity: from synapses to maps. *Annu. Rev. Neurosci.* **21**, 149–186 (1998).
- Gilbert, C. D. Adult cortical dynamics. *Physiol. Rev.* **78**, 467–485 (1998).
- Merzenich, M. M. *et al.* Progression of changes following median nerve section in the cortical representation of the hand in areas 3b and 1 in adult owl and squirrel monkeys. *Neuroscience* **10**, 639–665 (1983).
- Robertson, D. & Irvine, D. Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness. *J. Comp. Neurol.* **282**, 456–471 (1989).
- Kaas, J. H. *et al.* Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina. *Science* **248**, 229–231 (1990).
- Glazewski, S. & Fox, K. Time course of experience-dependent synaptic potentiation and depression in barrel cortex of adolescent rats. *J. Neurophysiol.* **75**, 1714–1729 (1996).
- Gilbert, C. D. & Wiesel, T. N. Receptive field dynamics in adult primary visual cortex. *Nature* **356**, 150–152 (1992).
- Fox, K. A critical period for experience-dependent synaptic plasticity in rat barrel cortex. *J. Neurosci.* **12**, 1826–1838 (1992).
- Diamond, M. E., Huang, W. & Ebner, F. F. Laminar comparison of somatosensory cortical plasticity. *Science* **265**, 1885–1888 (1994).
- Kirkwood, A., Rioult, M. C. & Bear, M. F. Experience-dependent modification of synaptic plasticity in visual cortex. *Nature* **381**, 526–528 (1996).
- Glazewski, S., Chen, C.-M., Silva, A. & Fox, K. Requirement for α -CaMKII in experience-dependent plasticity of the barrel cortex. *Science* **272**, 421–423 (1996).
- Garraghty, P. E. & Muja, N. NMDA receptors and plasticity in adult primate somatosensory cortex. *J. Comp. Neurol.* **367**, 319–326 (1996).
- Carvell, G. E. & Simons, D. J. Biometric analyses of vibrissal tactile discrimination in the rat. *J. Neurosci.* **10**, 2638–2648 (1990).
- Welker, C. & Woolsey, T. A. Structure of layer IV in the somatosensory neocortex of the rat: description and comparison with the mouse. *J. Comp. Neurol.* **158**, 437–454 (1974).
- Markram, H. & Tsodyks, M. Redistribution of synaptic efficacy between neocortical pyramidal cells. *Nature* **382**, 807–810 (1996).
- Abbot, L. F., Varela, J. A., Sen, K. & Nelson, S. B. Synaptic depression and cortical gain control. *Science* **275**, 220–224 (1997).
- Tsodyks, M. V. & Markram, H. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc. Natl Acad. Sci. USA* **94**, 719–723 (1997).
- McCasland, J. S. & Woolsey, T. A. High-resolution 2-deoxyglucose mapping of functional cortical columns in mouse barrel cortex. *J. Comp. Neurol.* **278**, 555–569 (1988).
- Armstrong-James, M., Diamond, M. E. & Ebner, F. F. An innocuous bias in whisker use in adult rats modifies receptive fields of barrel cortex neurons. *J. Neurosci.* **14**, 6978–6991 (1994).
- Darian-Smith, C. & Gilbert, C. D. Axonal sprouting accompanies functional reorganization in adult cat striate cortex. *Nature* **368**, 737–740 (1994).
- Florence, S. L., Taub, H. B. & Kaas, J. H. Large-scale sprouting of cortical connections after peripheral injury in adult macaque monkeys. *Science* **282**, 1117–1121 (1998).
- Mason, A., Nicoll, A. & Stratford, K. Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. *J. Neurosci.* **11**, 72–84 (1991).
- Nicolelis, M. A. L., Baccala, L. A., Lin, R. C. S. & Chapin, J. K. Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science* **268**, 1353–1358 (1995).
- Stuart, G. & Sakmann, B. Amplification of EPSPs by axosomatic sodium channels in neocortical pyramidal neurons. *Neuron* **15**, 1065–1076 (1995).
- Connors, B. W., Malenka, R. C. & Silva, L. R. Two inhibitory postsynaptic potentials, and GABA_A receptor-mediated responses in neocortex of rat and cat. *J. Physiol. (Lond.)* **406**, 443–468 (1988).
- Dobrunz, L. E. & Stevens, C. F. Heterogeneity of release probability, facilitation, and depletion at central synapses. *Neuron* **18**, 995–1008 (1997).
- Raastad, M., Storm, J. F. & Andersen, P. Putative single quantum and single fibre excitatory postsynaptic currents show similar amplitude range and variability in rat hippocampal slices. *Eur. J. Neurosci.* **4**, 113–117 (1992).
- Volgushev, M., Voronin, L. L., Chistiakova, M., Artola, A. & Singer, W. All-or-none excitatory postsynaptic potentials in the rat visual cortex. *Eur. J. Neurosci.* **7**, 1751–1760 (1995).
- Cash, S. & Yuste, R. Input summation by cultured pyramidal neurons is linear and position-independent. *J. Neurosci.* **18**, 10–15 (1998).

Acknowledgements. This work was supported by a Wellcome Trust Advanced Training Fellowship to G.T.F. and a grant to B.W.C. from NIH. We thank S. Patrick for technical assistance with histology; A. Akima for discussions on the optimal slicing angle; D. Pinto for checking the mathematics; and M. Bear, J. Gibson, M. Beierlein and D. Pinto for helpful comments on an earlier version of the manuscript. G.T.F. thanks R. Jones and D. Smith for encouragement with the current project.

Correspondence and requests for materials should be addressed to B.W.C. (e-mail: Barry_Cornors@brown.edu).

Selective inhibition of cocaine-seeking behaviour by a partial dopamine D₃ receptor agonist

Maria Pilla^{*†}, Sylvie Perachon^{*‡§}, François Sautel[¶], Fabrice Garrido^{||}, André Mann^{||}, Camille G. Wermuth^{||}, Jean-Charles Schwartz[‡], Barry J. Everitt[†] & Pierre Sokoloff[‡]

[†] Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

[‡] Unité de Neurobiologie et Pharmacologie Moléculaire, INSERM U 109, Centre Paul Broca, 2ter rue d'Alésia, 75014 Paris, France

[§] Laboratoire Bioprojet, 9 rue Rameau, 75002 Paris, France

^{||} Laboratoire de Pharmacochimie de la Communication Cellulaire CNRS

ERS 655, Faculté de Pharmacie, 74 route du Rhin BP 24, 67401 Illkirch, France

^{*} These authors contributed equally to this work.

Environmental stimuli that are reliably associated with the effects of many abused drugs, especially stimulants such as cocaine, can produce craving and relapse in abstinent human substance abusers^{1–4}. In animals, such cues can induce and maintain drug-seeking behaviour and also reinstate drug-seeking after extinction^{5–7}. Reducing the motivational effects of drug-related cues might therefore be useful in the treatment of addiction³. Converging pharmacological^{8,9}, human post-mortem¹⁰ and genetic¹¹ studies implicate the dopamine D₃ receptor¹² in drug addiction. Here we have designed BP 897, the first D₃-receptor-selective agonist, as assessed *in vitro* with recombinant receptors and *in vivo* with mice bearing disrupted D₃-receptor genes. BP 897 is a partial agonist *in vitro* and acts *in vivo* as either an agonist or an antagonist. We show that BP 897 inhibits cocaine-seeking behaviour that depends upon the presentation of drug-associated cues, without having any intrinsic, primary rewarding effects. Our data indicate that compounds like BP 897 could be used for reducing the drug craving and vulnerability to relapse that are elicited by drug-associated environmental stimuli.

Until now, it has proved difficult to identify or design a D₃

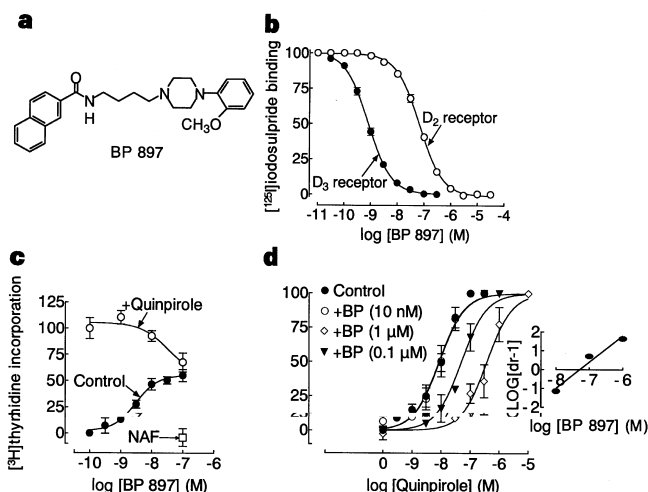


Figure 1 BP 897 is a potent and partial D₃ receptor agonist and a weak D₂ receptor antagonist. **a**, Chemical structure of BP 897. **b**, Inhibition by BP 897 of [¹²⁵I]iodosulpride binding to recombinant D₂ and D₃ receptors. **c**, In NG 108-15 cells expressing the D₃ receptor, BP 897 activated mitogenesis and this response was antagonized by the preferential D₃ receptor antagonist nafadotride (NAF, 1 μM). BP 897 also partially antagonized the response induced by quinpirole (10 nM). **d**, In CHO cells expressing the D₂ receptor, BP 897, while having no effect by itself, reversibly antagonized quinpirole-induced mitogenesis. The Schild plot (inset) allowed us to calculate a pK_i value for BP 897 of 7.29 (K_i = 51 nM).

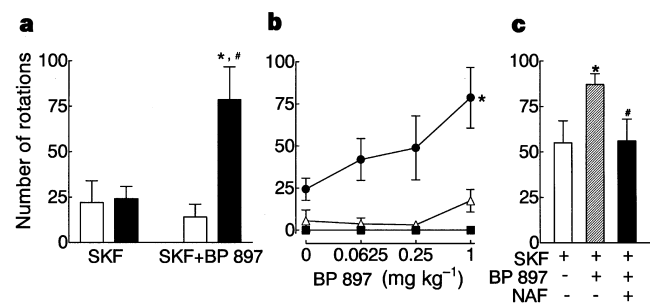


Figure 2 Agonistic effects of BP 897 on rotations in 6-OHDA-lesioned rats. **a**, Contralateral rotations induced by BP 897 (1 mg kg⁻¹) alone or in combination with SKF 38393 (SKF, 10 mg kg⁻¹), before (open bars) and after (filled bars) repeated twice daily administration of levodopa for 10 days ($F_{[5,45]} = 6.35$, $P = 0.0002$ by analysis of variance; * $P < 0.001$ vs. SKF after L-DOPA, # $P < 0.05$ vs. SKF + BP 897 before L-DOPA). Results are means $\pm$ s.e.m. of data obtained from 6–13 animals. **b**, BP 897 dose-dependently increased contralateral rotations induced by SKF 38393 (5 mg kg⁻¹). Squares, saline; triangles, 2.5 mg kg⁻¹ SKF; circles, 5 mg kg⁻¹ SKF ($F_{[3,39]} = 3.41$, $P = 0.027$; * $P < 0.05$ vs. SKF). **c**, The effect of BP 897 (1 mg kg⁻¹) on rotations induced by SKF (5 mg kg⁻¹) was antagonized by nafadotride (1 mg kg⁻¹) ($F_{[2,18]} = 9.20$, $P = 0.0018$ by analysis of variance, * $P < 0.01$ vs. SKF and # $P < 0.01$ vs. SKF + BP 897).

receptor agonist that has no agonist activity at the closely related D₂ receptor¹³. By screening a series of newly designed molecules for their differential affinity at recombinant human dopamine D₂ and D₃ receptors, we selected BP 897 (Fig. 1a), which displayed a high affinity at the D₃ receptor ($K_i = 0.92 \pm 0.2$ nM, mean $\pm$ s.e.m., $n = 12$) and a 70 times lower affinity at the D₂ receptor ($K_i = 61 \pm 0.2$ nM, mean $\pm$ s.e.m., $n = 8$; Fig. 1b). BP 897 also displayed low affinities at D₁ and D₄ receptors ($K_i = 3$ and 0.3 μ M, respectively), as well as at α_1 and α_2 adrenergic receptors ($K_i = 60$ and 83 nM, respectively), 5HT_{1A} and 5HT₇ receptors ($K_i = 84$ and 345 nM, respectively), and negligible affinities ($K_i > 1$ μ M) at muscarinic, histamine and opiate receptors (not shown).

In NG 108-15 cells expressing the human D₃ receptor, BP 897 inhibited forskolin-induced cyclic AMP accumulation with an EC₅₀ of 1.0 ± 0.3 nM, although to a maximal extent ($-59\% \pm 4\%$) lower than that elicited by dopamine or the full agonist quinpirole ($-82 \pm 6\%$). This response was completely reversed by 1 μ M haloperidol, a D₂/D₃ receptor antagonist (data not shown). In these cells, BP 897 also increased mitogenesis, another D₃-receptor-mediated response¹⁴ (EC₅₀ = 3 ± 1 nM), which at maximum was only 55% of that elicited by quinpirole and was antagonized by nafadotride, a preferential D₃ receptor antagonist¹⁵ (Fig. 1c). BP 897 also antagonized quinpirole-induced mitogenesis, up to the maximal level reached with BP 897 alone (Fig. 1c).

In contrast, in cells expressing the D₂ receptor, BP 897 (1 μ M), unlike quinpirole, did not inhibit cyclic AMP accumulation or trigger mitogenesis; it reversibly antagonized quinpirole-induced mitogenesis, but only at concentrations much larger than those required to stimulate the D₃ receptor (Fig. 1d). Hence, BP 897 is a selective, potent, but partial (intrinsic activity ~ 0.6) D₃ receptor agonist and a weak D₂ receptor antagonist *in vitro*.

Initial *in vivo* binding experiments in mouse striatum using [³H]N-propylnorapomorphine showed that D₂-receptor occupancy is achieved by BP 897 with an ED₅₀ of ~ 15 mg kg⁻¹ and maximally maintained for 2 h. The absence of dopamine agonist activity at the D₂ receptor was shown by the inability of BP 897 to induce stereotypies (J. Costentin, personal communication). BP 897 produced catalepsy in rats, a classical response to D₂-receptor blockade, with an ED₅₀ of ~ 12 mg kg⁻¹. Taking into account the respective affinities and potencies of BP 897 at D₂ and D₃ receptors, this indicated that D₃-receptor occupancy was to be expected at an ED₅₀ below 0.5 mg kg⁻¹.

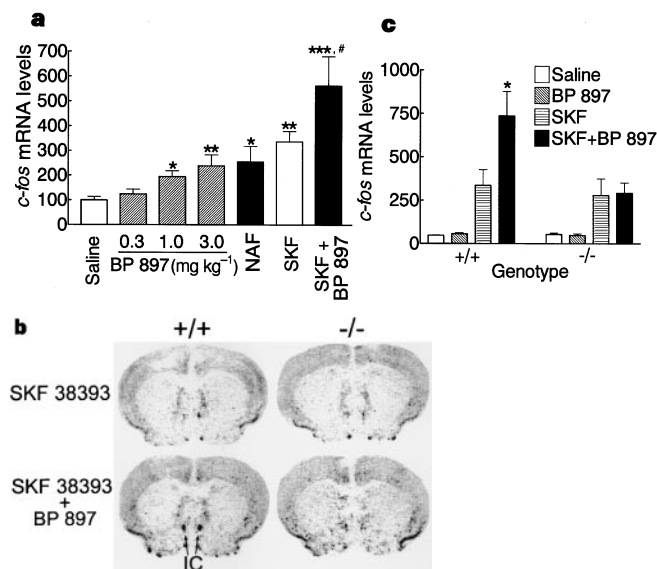


Figure 3 Antagonistic effects of BP 897 on *c-fos* gene expression in the islands of Calleja. **a**, Effect of BP 897 (0.3–1 mg kg⁻¹), nafadotride (NAF, 1 mg kg⁻¹), SKF 38393 (SKF, 10 mg kg⁻¹) or the combination of SKF (10 mg kg⁻¹) and BP 897 (1 mg kg⁻¹) on the expression of *c-fos* mRNA in rats. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ vs. saline; # $P < 0.05$ vs. SKF by analysis of variance. **b**, *In situ* hybridization signals of *c-fos* mRNA in wild-type (+/+) or D₃ receptor knockout mice (-/-) treated with SKF 38393 (10 mg kg⁻¹) or the combination of SKF and BP 897 (1 mg kg⁻¹). **c**, Pictures as in **b** were analysed from 5–10 animals. Two-way analysis of variance indicates a significant effect of genotype ($F_{[1,55]} = 4.17$, $P = 0.04$) and a significant treatment $\times$ genotype interaction ($F_{[3,55]} = 3.48$, $P = 0.02$). * $P < 0.001$ vs. all other groups.

We have assessed further the agonist/antagonist potency of BP 897 *in vivo* on two D₃-receptor-mediated responses. In hemiparkinsonian rats, rotations contralateral to the lesion are elicited by stimulation of hypersensitive postsynaptic dopamine receptors in the dopamine-denervated striatum. Repeated pretreatment with levodopa (L-DOPA) sensitizes the animals to this drug, an effect attributable to the induction of D₃ receptor gene expression in neurons containing D₁ receptors¹⁶. Thus, monitoring rotations in these animals provides a reliable behavioural model of D₃-receptor stimulation. In such animals, BP 897 potentiated rotations elicited by the D₁-receptor-selective agonist SKF 38393 in a dose-dependent manner (Fig. 2a, b). This potentiation did not occur before L-DOPA treatment, that is, before induction of D₃-receptor expression, and was abolished by co-treatment with nafadotride used at a dose compatible with a selective D₃-receptor blockade¹⁶ (Fig. 2a, c).

In a second *in vivo* model, we measured *c-fos* expression in granule cells of the islands of Calleja, co-expressing D₁ and D₃ receptors, which mediate tonic but opposite effects on *c-fos* expression. For instance, D₁ receptor agonists and antagonists enhance and impair *c-fos* expression, respectively, whereas corresponding D₃ receptor agents (also active at the D₂ receptor) exert the opposite actions¹⁷. In the islands of Calleja of rats, BP 897 enhanced *c-fos* messenger RNA with an ED₅₀ of ~ 0.3 mg kg⁻¹, an effect similar in direction and amplitude to that of nafadotride (Fig. 3a); in addition, in contrast to D₂/D₃ receptor agonists¹⁷, BP 897 potentiated the response to a D₁ receptor agonist (Fig. 3a).

We used mice bearing mutations invalidating the D₃-receptor gene¹⁸ to assess the specificity of the effects of BP 897 on *c-fos* gene expression. In mice (wild-type or mutant), BP 897 had no marked effects on *c-fos* mRNA (Fig. 3c). However, BP 897 strongly potentiated the effects of the D₁ receptor agonist in wild-type mice (Fig. 3b, c), an effect that was not seen in mice lacking the D₃ receptor. BP 897 shows extreme region-related specificity: it increased *c-fos* mRNA levels in wild-type mice only in the islands

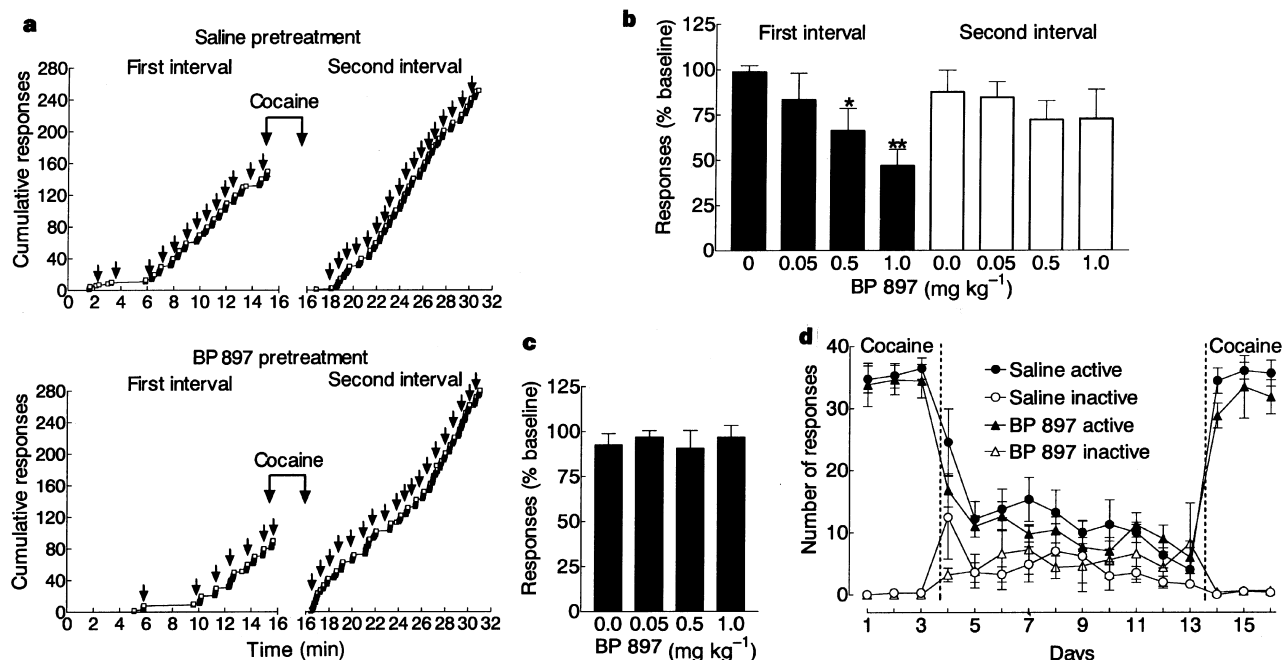


Figure 4 BP 897 inhibits cue-controlled cocaine-seeking but has no reinforcing effects. **a**, Cumulative response records for a representative rat receiving a pre-injection of saline (top) or BP 897 0.5 mg kg⁻¹ (bottom) on responding during the first or second interval of a cocaine self-administration session under a FI15(FR10:S) second-order schedule of reinforcement (see Methods). The arrows indicate contingent, 1 s presentations of the CS+, following the completion of each fixed ratio of 10 responses on the active lever. In the second interval, but not the first, responding is under the influence of the self-administered cocaine. Pretreatment with BP 897 markedly reduces responding in the first, but not the second, interval. **b**, Dose-related effects of BP 897 on cocaine-seeking behaviour evaluated as in **a** ($n = 11$). BP 897 decreased responding for cocaine during the first, cocaine-free, interval of a second-order schedule ($F[3, 30] = 6.45$, $P = 0.01$ by analysis of variance, $*P < 0.05$ and $**P < 0.01$ vs. saline), but not during the second interval ($F[3, 30] = 0.93$). **c**, BP 897 has no significant effect on the self-administration of cocaine under a continuous reinforcement schedule

($F[3, 27] = 0.21$). **d**, Effects of replacing cocaine with BP 897 (0.5 mg ml⁻¹, 100 μ l per infusion: 0.05 mg i.v.) or saline (100 μ l per infusion) on i.v. cocaine self-administration. After 2 days of replacement, responding on the active lever in all subjects had decreased to levels not significantly different from the inactive lever, that is, self-administration behaviour had extinguished. There was no difference between the effects of replacement with BP 897 or saline. Re-substituting cocaine for BP 897 or saline resulted in the prompt reinstatement of drug self-administration in both groups. Analysis of variance revealed no effect of treatment (replacement with BP 897 or saline, $F[1, 10] = 0.37$, $P > 0.5$); a main effect of lever ($F[1, 10] = 219.12$, $P < 0.0001$) reflecting more responses on the active than the inactive lever; a main effect of day ($F[15, 150] = 18.91$, $P < 0.0001$); and a lever $\times$ day interaction ($F[15, 150] = 50.90$, $P < 0.001$) that reflected the decrease in responding on the active lever in both groups following BP 897 or saline substitution on day 3 and on cocaine re-substitution on day 14. There were no significant interactions of treatment with either lever or day.

of Calleja.

Thus, *in vivo*, BP 897 increases D₁-receptor-mediated responses by acting as either an agonist or an antagonist, depending upon the response considered, consistent with its partial agonist properties *in vitro*. This dual activity may be explained by various models of receptor activation, in which the efficacy of a ligand depends not only on its intrinsic property, but also on receptor–G-protein coupling efficiency^{19,20}, which may vary between tissues. The direction of the response may also depend on the level of the endogenous transmitter. Accordingly, BP 897 acts as a receptor agonist on rotations elicited in the dopamine-depleted brain and as a receptor antagonist on *c-fos* expression maintained by a dopaminergic tone.

Taken in the context of the importance of the D₃ receptor in mediating the reinforcing effects of cocaine^{8,21}, the theoretical use of partial dopamine receptor agonists as treatments for addictive behaviour²² and the impact of drug-associated cues on drug craving and relapse in humans^{2,5,23,24}, we have investigated the effects of BP 897 in a model of cocaine-seeking behaviour in which drug cues have a demonstrably important controlling role^{7,24,25}.

Rats were trained to self-administer cocaine under a second-order schedule of reinforcement^{7,24,25}, responses being reinforced by presentation of a light stimulus paired with each cocaine infusion. The light-conditioned stimulus, acting as a conditioned reinforcer, maintains the lever-pressing response for a protracted period (usually a fixed interval of 15 min; FI15) before cocaine infusion (Fig. 4a, top). This drug-seeking behaviour⁷ occurs in the absence of any pharmacological effects of cocaine (as the animals are never

'primed') and depends critically upon the contingent presentation of the cocaine cue; its omission greatly reduces responding, even though intravenous cocaine is still available⁷. Administration of BP 897 before testing reduced cocaine-seeking behaviour before the first infusion of cocaine, in a dose-dependent manner, at doses similar to those at which BP 897 produced its responses on rotations and *c-fos* expression (Fig. 4a, bottom, and 4b). Importantly, the effect of BP 897 occurred without disrupting the pattern of responding and without increasing the latency to initiate responding at any dose up to 1 mg kg⁻¹ (data not shown), indicating the absence of any nonspecific, for example motor, effects that might have disrupted performance. Preliminary data indicate that the reduction in cocaine-seeking behaviour following administration of BP 897 is prevented by pretreatment with nafadotride (M.P., J.-C.S., P.S. and B.J.E., unpublished observations).

Under the second-order schedule used here, self-administered cocaine increases subsequent drug-seeking behaviour⁷, but BP 897 had no effect on this cocaine-induced increase in responding in the second interval (Fig. 4a, b), at a time when BP 897 is still active according to our *in vivo* binding experiments (see above). These results indicate a dissociation in the effects of a partial D₃ receptor agonist on drug-cue-controlled cocaine-seeking behaviour and on the reinforcing, including response-rate-increasing, effects of cocaine, as the effects of BP 897 were only apparent in the first (cocaine-unaffected) interval of a session.

To investigate this dissociation more specifically, we first tested whether BP 897 pretreatment could modify intravenous cocaine

self-administration under a continuous reinforcement schedule, and found that it had no effect (Fig. 4c). Next, BP 897 (at an intravenous dose equivalent to the highest effective dose when given intraperitoneally) or saline were substituted for cocaine to determine whether the drug possessed reinforcing properties sufficient to maintain drug-taking in animals with a daily history of intravenous cocaine self-administration. Responding declined identically over days in rats in which either BP 897 or saline had been substituted for cocaine (Fig. 4d). Thus, from three sets of independent observations, BP 897 appears to have no intrinsic reinforcing effects, but can nevertheless reduce significantly conditioned cue-controlled cocaine-seeking behaviour.

The ability of BP 897 to reduce drug-seeking behaviour in the absence of any intrinsic reinforcing properties, although unprecedented, is consistent with data indicating that dissociable neural mechanisms underlie responding with conditioned reinforcement and responding for cocaine itself^{6,25,26}. However, these dissociable mechanisms clearly interact and projections from the basolateral amygdala to the nucleus accumbens may provide the neuroanatomical basis of this interaction, with the accumbens shell region (where neurons express high D₃ receptor levels²⁷) providing a site for dopaminergic modulation, for example by the D₃-receptor-selective partial agonist used here. The partial agonistic character and D₃-receptor selectivity of BP 897 seem to be essential for its dissociated actions, as the full and non-selective agonist quinpirole, in addition to cocaine itself, also reduces cue-controlled cocaine-seeking behaviour (A. Markou, M. Arroyo & B.J.E., unpublished results), whereas, in marked contrast to BP 897, it possesses intrinsic rewarding properties, being both self-administered and also able to enhance the reinforcing effects of cocaine⁸. Furthermore, PNU99194A, a D₃ receptor antagonist, has no effect on drug-seeking behaviour (M.P., M. Arroyo & B.J.E., unpublished results). The absence of effects of BP 897 on cocaine reinforcement may also be attributable to its partial agonism, as full agonists and antagonists have opposite effects⁹. Partial agonism of BP 897 may confer a 'buffering' capacity, allowing it to oppose, as an antagonist, a chain of neural events initiated by a conditioned increase in dopamine release^{28,29}, while maintaining, as an agonist, a moderate degree of D₃-receptor stimulation. Finally, as D₁ receptor-mediated mechanisms seem to be important in a model of reinstatement of cocaine self-administration after extinction³⁰, BP 897 may exert part of its actions by potentiating some D₁-receptor-mediated effects in the restricted subset of neurons in which D₁ and D₃ receptors co-localize¹⁷. □

Methods

Drug screening. The synthesis of BP 897 will be described elsewhere. We performed binding studies at D₁, D₂, D₃ and D₄ receptors using Chinese hamster ovary (CHO) cells expressing recombinant human receptors and [³H]SCH23390 (for D₁), [¹²⁵I]iodosulpride (for D₂ and D₃) and [³H]spiperone (for D₄) as described²⁷. Additional receptor screening was performed by Panlabs Taiwan Ltd.

Measurement of cAMP and mitogenesis. We used NG 108-15 cells expressing the D₃ receptor or CHO cells expressing the D₂ receptor¹⁴. cAMP levels were measured after a 10 min stimulation by forskolin (1 µM) and drugs in the presence of isobutylmethylxanthine (10 µM) using the Rianen [¹²⁵I]cAMP radioimmunoassay kit (DuPont NEN). Mitogenesis was measured by incorporation of [³H]thymidine during a 2 h pulse following an overnight incubation with drugs¹⁴.

Measurement of striatal dopamine receptor occupancy and catalepsy. We measured striatal dopamine receptor occupancy in mice killed 15 min after an intravenous (i.v.) injection of [³H]N-propyl-norapomorphine and 30 min after an intraperitoneal (i.p.) injection of saline or BP 897 (1–30 mg kg⁻¹). For assessment of catalepsy, rats were placed, 30 min after an i.p. injection of BP 897 (1–30 mg kg⁻¹), in a device in which the forepaws were held 10 cm above the floor with a horizontal bar. Dose–response curves were generated from the percentage of animals holding the position for at least 5 s.

Rotational behaviour. Anaesthetized Wistar male rats were infused with

hydroxydopamine (8 µg in 4 µl of 0.05% ascorbic acid in saline) in the medial forebrain bundle¹⁶. Four weeks later, we measured rotations for 5 min, starting 35 min after administration of drugs after a 5-min period of habituation in vertical Perspex cylinders (30 cm diameter). Rotations were measured again after animals had received, twice a day for 10 days, levodopa (as L-DOPA-methylester, 50 mg kg⁻¹ i.p., in combination with benserazide, 12.5 mg kg⁻¹ i.p.).

In situ hybridization. Rats were habituated to handling and injections by repeated intraperitoneal administration of saline, during four days before the experiments. They were killed 30 min after administration of the drugs and their brains frozen in isopentane maintained at –30 °C. Slices (10 µm) were hybridized with a ³³P-labelled c-fos antisense RNA probe as described¹⁷. We analysed autoradiograms with an image analyser, the islands of Calleja being visualized after superimposition of images obtained after hemalun counterstaining. Similar experiments were performed on 8 µm slices from wild-type and D₃ receptor knock-out mice (generations F3 and F4), created by S. Fuchs and D. Accili (Weizmann Institute, Rehovot, Israel)¹⁸ and initially bred in France by W. Rostene and C. Betancur (INSERM U336, Paris).

Cocaine self-administration under continuous and second-order schedules of reinforcement. Male Listar hooded rats (Charles River) were implanted with indwelling intravenous catheters as described⁷. Each rat was placed in an operant chamber where two retractable levers were available and trained to self-administer cocaine under a continuous reinforcement schedule (a fixed ratio 1 of responding: FR1). Depression of the active lever resulted in an infusion of 100 µl of a solution of cocaine hydrochloride (2.5 mg ml⁻¹, dissolved in 0.9% saline; 0.25 mg per infusion), extinction of the house light, illumination of a stimulus light (the conditioned stimulus; CS+) above the active lever for 20 s, and then retraction of both levers from the chamber. Depression of the inactive lever had no programmed consequence and was also followed by retraction of both levers. Active and inactive levers were counter-balanced left and right across subjects. Cocaine self-administration under continuous reinforcement was used to evaluate the effects of BP 897 (administered i.p. 30 min before the session) on cocaine self-administration and when substituted for cocaine.

When rats had acquired a stable pattern of responding (three days of constant responding ±10%), we introduced a second-order schedule of the type FRx(FRy:S), as described⁷. Progressively, the value of x was increased to 10 (that is, each active lever response produced a CS+, and 10 such responses were followed by a drug infusion and a 1 s illumination of the light CS+). In a second stage, the value of y was also increased progressively to 10, so that 10 responses were required to produce the CS+, and 10 such units (100 responses; 10 CS+ presentations) resulted in a cocaine infusion (and a 20 s presentation of the CS+). Ultimately, the form of the second-order schedule in operation was an overall fixed interval (FI) 15 min (i.e. FI15(FR10:S)) in which every FR10 response resulted in presentation of the CS+ as before, but the cocaine infusion was delivered only on completion of the first FR10 responses after the 15 min fixed interval had timed out. Responding for the first cocaine infusion in a daily session, evaluated 30 min after intraperitoneal administration of saline or BP 897, therefore occurred in the absence of any effects of cocaine and represents drug-cue-controlled drug-seeking behaviour. Following the first cocaine infusion, responding in a second 15 min interval is affected by the rewarding and other (for example, response-rate-enhancing) effects of cocaine.

In substitution experiments, rats first acquired i.v. cocaine self-administration under continuous reinforcement. When stable baseline rates of cocaine self-administration were established (3 consecutive days with same number of responses ±10%), BP 897 (0.5 mg kg⁻¹) or saline were substituted for cocaine and self-administration was allowed to continue for a further 10 days. At this point, cocaine at the training dose was re-substituted for BP 897 or saline.

Received 4 February; accepted 4 May 1999.

- Wikler, A. Dynamics of drug dependence. *Arch. Gen. Psychiatry* **28**, 611–616 (1973).
- O'Brien, C. P., Childress, A. R., McMellan, A. T. & Ehrman, R. A. A learning model of addiction. *Res. Publ. Assoc. Res. Nerv. Ment. Dis.* **70**, 157–177 (1992).
- O'Brien, C. P. & McMellan, A. T. Myths about the treatment of addiction. *Lancet* **347**, 237–240 (1996).
- Grech, D. M., Speelman, R. D. & Bergman, J. Self-administration of D₁ receptor agonists by squirrel monkeys. *Psychopharmacology* **125**, 97–104 (1996).
- de Wit, H. & Stewart, J. Reinstatement of cocaine-reinforced responding in the rat. *Psychopharmacology* **75**, 134–143 (1981).
- Meil, W. M. & See, R. E. Lesions of the basolateral amygdala abolish the ability of drug associated cues to reinstate responding during withdrawal from self-administered cocaine. *Behav. Brain Res.* **87**, 139–

- 148 (1997).
7. Arroyo, M., Markou, A., Robbins, T. W. & Everitt, B. J. Acquisition, maintenance and reinstatement of intravenous cocaine self-administration under a second-order schedule of reinforcement in rats: effects of conditioned cues and continuous access to cocaine. *Psychopharmacology* **140**, 331–344 (1999).
 8. Caine, S. B. & Koob, G. F. Modulation of cocaine self-administration in the rat through D₃ dopamine receptors. *Science* **260**, 1814–1816 (1993).
 9. Caine, S. B. *et al.* D₃ receptor test *in vitro* predicts decreased cocaine self-administration in rats. *Neuroreport* **8**, 2373–2377 (1997).
 10. Staley, J. K. & Mash, D. C. Adaptive increase in D₃ dopamine receptors in the brain reward circuits of human cocaine fatalities. *J. Neurosci.* **16**, 6100–6106 (1996).
 11. Duaux, E. *et al.* Homozygosity at the dopamine D₃ receptor gene is associated with opioid dependence. *Mol. Psychiatry* **3**, 333–336 (1998).
 12. Sokoloff, P., Giros, B., Martres, M.-P., Bouthenet, M.-L. & Schwartz, J.-C. Molecular cloning and characterization of a novel dopamine receptor (D₃) as a target for neuroleptics. *Nature* **347**, 146–151 (1990).
 13. Xu, M., Koeltzow, T. E., Cooper, D. C., Tonegawa, S. & White, F. J. Dopamine D₃ receptor mutant and wild-type mice exhibit identical responses to putative D₃ receptor-selective agonists and antagonists. *Synapse* **31**, 210–215 (1999).
 14. Pilon, C. *et al.* Functional coupling of the human dopamine D₃ receptor in a transfected NG 108-15 neuroblastoma-glioma hybrid cell line. *Eur. J. Pharmacol.* **268**, 129–139 (1994).
 15. Sautel, F. *et al.* Nafadotride, a potent preferential dopamine D₃ receptor antagonist, activates locomotion in rodents. *J. Pharmacol. Exp. Ther.* **275**, 1239–1246 (1995).
 16. Bordet, R. *et al.* Induction of dopamine D₃ receptor expression as a mechanism of behavioral sensitization to levodopa. *Proc. Natl Acad. Sci. USA* **94**, 3363–3367 (1997).
 17. Ridray, S. *et al.* Coexpression of dopamine D₁ and D₃ receptors in rat ventral striatum: opposite and synergistic functional interactions. *Eur. J. Neurosci.* **10**, 1676–1686 (1998).
 18. Accili, D. *et al.* A targeted mutation of the D₃ receptor gene is associated with hyperactivity in mice. *Proc. Natl Acad. Sci. USA* **93**, 1945–1949 (1996).
 19. Costa, T., Ogino, Y., Munson, P. J., Oraran, H. O. & Rodbard, D. Drug efficacy at guanine nucleotide-binding regulatory protein linked receptors: thermodynamic interpretation of negative antagonism and of receptor activity in the absence of ligand. *Mol. Pharmacol.* **41**, 549–560 (1992).
 20. Leff, P. The two-state model of receptor activation. *Trends Pharmacol. Sci.* **16**, 89–97 (1995).
 21. Caine, S. B. & Koob, G. F. Pretreatment with the dopamine agonist 7-OH-DPAT shifts the cocaine self-administration dose-effect function to the left under different schedules in the rat. *Behav. Pharmacol.* **6**, 333–347 (1995).
 22. Pulvirenti, L. & Koob, G. F. Dopamine agonists, partial agonists and psychostimulant addiction. *Trends Pharmacol. Sci.* **15**, 374–379 (1994).
 23. Grant, S. *et al.* Activation of memory circuits during cue-elicited cocaine craving. *Proc. Natl Acad. Sci. USA* **93**, 12040–12045 (1996).
 24. Weissenborn, R., Robbins, T. W. & Everitt, B. J. Effects of medial prefrontal or anterior cingulate cortex lesions on responding for cocaine under fixed-ratio and second-order schedules of reinforcement in rats. *Psychopharmacology* **134**, 242–257 (1997).
 25. Whitelaw, R. B., Markou, A., Robbins, T. W. & Everitt, B. J. Excitotoxic lesions of the basolateral amygdala impair the acquisition of cocaine-seeking behaviour under a second-order schedule of reinforcement. *Psychopharmacology* **127**, 213–224 (1996).
 26. Everitt, B. J. & Robbins, T. W. In *The Amygdala: Neurological Aspects of Emotion, Memory and Mental Dysfunction* (ed. Aggleton, J.) 401–429 (Wiley, New York, 1992).
 27. Lévesque, D. *et al.* Identification, characterization and localization of the dopamine D₃ receptor in rat brain using 7-[³H]-hydroxy-N,N di-n-propyl-2-aminotetralin. *Proc. Natl Acad. Sci. USA* **89**, 8155–8159 (1992).
 28. Fontana, D. J., Post, R. M. & Pert, A. Conditioned increases in mesolimbic dopamine overflow by stimuli associated with cocaine. *Brain Res.* **629**, 31–39 (1993).
 29. Di Ciano, P., Blaha, C. D. & Phillips, A. G. Conditioned changes in dopamine oxidation currents in the nucleus accumbens of rats by stimuli paired with self-administration or yoked-administration of d-amphetamine. *Eur. J. Neurosci.* **10**, 1121–1127 (1998).
 30. Self, D. W., Barnhart, W. J., Lehman, D. A. & Nestler, E. J. Opposite modulation of cocaine-seeking behavior by D₁-like and D₂-like dopamine-receptor agonists. *Science* **271**, 1586–1589 (1996).

Acknowledgements. We thank C. Pilon for technical assistance and S. Fuchs, W. Rostene and C. Betancur for providing the D₃-receptor-deficient mice. The work was supported by a Biomed II Programme grant to P.S., C.G.W. and B.J.E., a National Institute on Drug Abuse grant to P.S. and C.G.W. and an MRC Programme grant to B.J.E.

Correspondence and requests for materials should be addressed to P.S. (e-mail: sokol@broca.inserm.fr) or B.J.E. (e-mail: bje10@cus.cam.ac.uk).

Circadian rhythms in olfactory responses of *Drosophila melanogaster*

Balaji Krishnan, Stuart E. Dryer & Paul E. Hardin

Department of Biology and Biochemistry and Biological Clocks Program, University of Houston, Houston, Texas 77204-5513, USA

The core mechanism of circadian timekeeping in arthropods and vertebrates consists of feedback loops involving several clock genes, including *period* (*per*) and *timeless* (*tim*)^{1,2}. In the fruitfly *Drosophila*, circadian oscillations in *per* expression occur in chemosensory cells of the antennae, even when the antennae are excised and maintained in isolated organ culture³. Here we demonstrate a robust circadian rhythm in *Drosophila* in electro-

physiological responses to two classes of olfactory stimuli. These rhythms are observed in wild-type flies during light–dark cycles and in constant darkness, but are abolished in *per* or *tim* null-mutant flies (*per*⁰¹ and *tim*⁰¹) which lack rhythms in adult emergence and locomotor behaviour. Olfactory rhythms are also abolished in the *per* 7.2:2 transgenic line in which *per* expression is restricted to the lateral neurons of the optic lobe⁴. Because *per* 7.2:2 flies do not express *per* in peripheral oscillators, our results provide evidence that peripheral circadian oscillators are necessary for circadian rhythms in olfactory responses. As olfaction is essential for food acquisition, social interactions and predator avoidance in many animals, circadian regulation of olfactory systems could have profound effects on the behaviour of organisms that rely on this sensory modality.

We determined olfactory rhythms by measuring electroantennogram (EAG) responses to odors^{5,6}. Typical EAG responses to the food odorant ethyl acetate (Fig. 1) were robust, dose-dependent and highly reproducible, both within and across preparations. The EAG responses did not saturate over the range of ethyl acetate concentrations used in these experiments (Fig. 1b), as described in studies of *Drosophila* by others⁷. Some desensitization was observed during odorant application (over a period of several seconds; Fig. 1c), but recovery from desensitization was invariably complete when odorant pulses were delivered at 30 s intervals (Fig. 1d). Because it was not possible to measure EAG amplitudes continuously from a single fly for long periods of time, we obtained mean EAG amplitudes from groups of animals at different times of day. Complete dilution–response curves were obtained for every animal.

To determine whether diurnal rhythms were present, recordings were made from flies maintained in light–dark (LD) 12:12 cycles. Recordings in the dark were made under infrared optics. Figure 2a shows that mean EAG responses to ethyl acetate were elevated during the middle of the night in wild-type flies during LD cycles, but not in *per*⁰¹ or *tim*⁰¹ mutant flies. Data shown in this and subsequent figures are based on mean responses to the highest concentration of ethyl acetate tested (a dilution of 1:10⁴). However, similar temporal patterns were observed at lower odorant concentrations (data not shown). Analyses of these data by two-way ANOVA indicate statistically significant ($P < 0.0001$) effects of time of day, genotype and their interaction. The significant interaction effect indicates that genotype alters the daily rhythm in mean EAG amplitude. *Post hoc* analyses indicate that mean EAG responses observed at Zeitgeber time (ZT) 13–17 were significantly ($P < 0.005$) greater than those observed at ZT21, ZT1, ZT5 or ZT9 in wild-type flies. In contrast, we observed no statistically significant changes in mean EAG responses to ethyl acetate at different times of day in *per*⁰¹ or *tim*⁰¹ mutant flies in LD (Fig. 2a).

To determine whether these rhythms are endogenously controlled, we made EAG recordings from flies on day 2 of constant darkness (DD) after 3 days' entrainment in LD. A similar temporal pattern of olfactory sensitivity was observed, indicating circadian control (Fig. 2b). The overall effects of time of day, genotype and their interactions were statistically significant ($P < 0.0001$). Mean EAG responses in wild-type flies increased gradually between circadian time (CT)9 and CT13, and peaked during the middle of the subjective night (CT17). EAG responses returned to basal levels at CT21 and CT1. Mean EAG responses observed at CT13–17 were significantly ($P < 0.005$) greater than those observed at CT21, CT1, CT5 or CT9. In contrast, there was no significant change in mean EAG responses as a function of the time of day in *per*⁰¹ or *tim*⁰¹ mutant flies in DD (Fig. 2b). A similar pattern was observed in olfactory responses to benzaldehyde (Fig. 2c), an odorant that, in contrast to ethyl acetate, causes behavioural avoidance in *Drosophila*⁸. The mean amplitude of EAG responses to benzaldehyde (dilution of 1:10³) was smaller than those evoked by ethyl acetate, as described by others⁷, possibly owing to the lower vapour pressure of this odorant. However, the phase of the rhythm was

similar, with peak responses at CT17. The circadian rhythm in EAG responses to benzaldehyde in DD was abolished in *tim*⁰¹ flies, and the effects of time of day, genotype and their interaction were statistically significant ($P < 0.0001$). This indicates that circadian rhythms in olfactory responses are not limited to food odours.

Previous studies have shown that circadian rhythms in *per* expression in wild-type *Drosophila* persist for at least 1 day in LL, but dampen to arrhythmicity by day 4 of LL^{9,10}. We observed a

circadian rhythm in olfactory responses to ethyl acetate in wild-type but not *per*⁰¹ or *tim*⁰¹ flies examined on day 1 of LL (Fig. 3a). The effects of time of day, genotype and their interactions were statistically significant ($P < 0.001$) on day 1 of LL. Circadian rhythms in EAG responses to ethyl acetate were fully damped by day 4 of LL. Two-way ANOVA and *post hoc* analyses indicate no significant effects of time of day on mean EAG responses in any genotype (Fig. 3b). However, a small but statistically significant ($P < 0.001$)

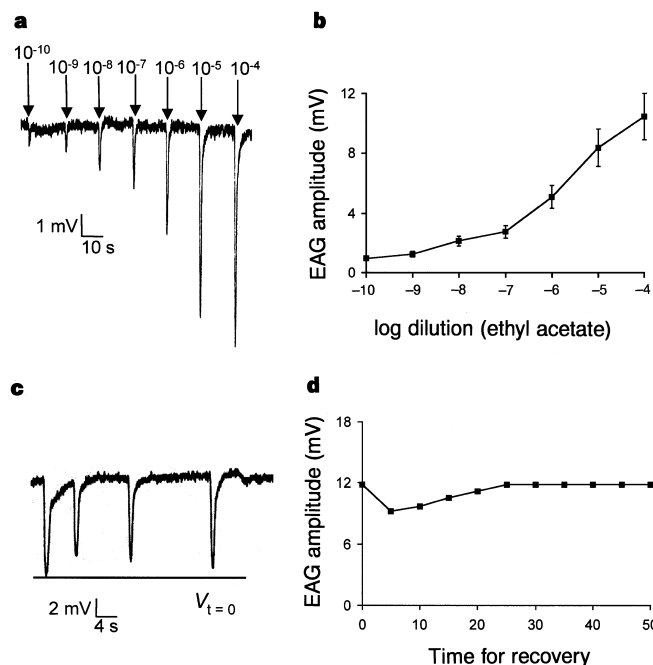


Figure 1 Characteristics of EAG responses to ethyl acetate in wild-type (Canton-S) female *Drosophila melanogaster*. **a**, Typical EAG responses to different dilutions of ethyl acetate (indicated above trace). **b**, Summary of dose-dependence of peak EAG responses to ethyl acetate in 24 wild-type flies. Data are represented as mean $\pm$ s.e.m. **c**, Recordings showing typical desensitization

and recovery from desensitization of responses to ethyl acetate ($1:10^4$ dilution in paraffin). **d**, Summary of data on recovery from desensitization in the same fly shown in **c**. Ordinate of graph represents interval between subsequent applications of $1:10^4$ ethyl acetate. Note complete recovery from desensitization with 30-s intervals between odorant pulses.

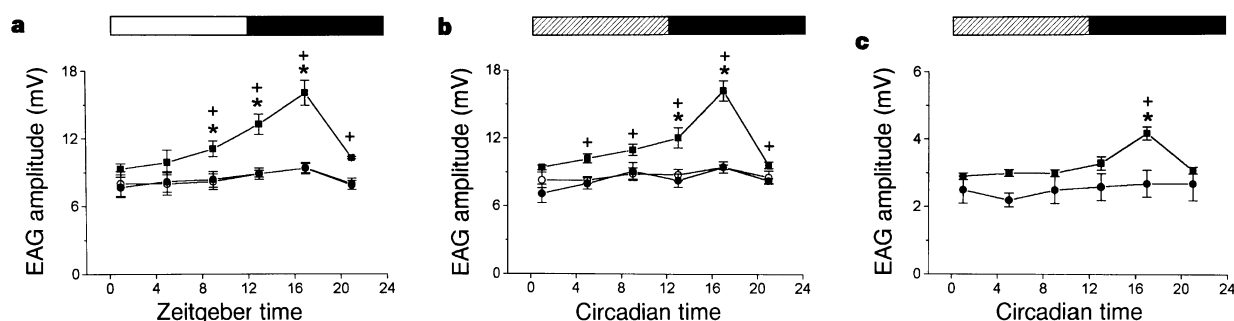


Figure 2 Olfactory responses at different times of day in different strains of *D. melanogaster*. Each point represents mean odorant-evoked EAG responses in the specified strain. Error bars denote s.e.m. **a**, Diurnal changes in mean EAG responses to ethyl acetate (dilution of $1:10^4$) on day 4 of LD 12:12 cycles. Each point represents the mean response of 24 flies. The white and black bars above the graph denotes when lights were on and off, respectively. Note EAG responses of wild-type flies (filled squares) increase during ZT9–13, peak at ZT17 and return to baseline thereafter. This rhythm was not observed in *per*⁰¹ (open circles) or *tim*⁰¹ (filled circles) flies. Overall effects of time of day ($F_{T5,414} = 48.67$), genotype ($F_{G2,414} = 215.13$) and their interaction ($F_{TG10,414} = 14.01$) are statistically significant ($P < 0.0001$) by two-way ANOVA. Asterisks denote significant ($P < 0.05$) increase in EAG responses in wild-type flies at ZT13 and ZT17 compared with responses at all other times of day in wild-type flies. Crosses indicate significant increase in EAG responses in wild-type flies compared with *per*⁰¹ and *tim*⁰¹ flies at the same

times of day. *Post hoc* analyses indicate no significant differences as a function of time of day in *per*⁰¹ or *tim*⁰¹ flies. **b**, A similar pattern in EAG responses to ethyl acetate is observed in flies free-running on day 2 of DD, indicating circadian control. Each point represents mean response in 24 flies. Hatched portion of bar above graph indicates subjective day, while black portion of bar indicates subjective night. Overall effects of time of day ($F_{T5,414} = 69.17$), genotype ($F_{G2,414} = 239.81$) and their interaction ($F_{TG10,414} = 19.09$) are statistically significant ($P < 0.0001$) by two-way ANOVA. **c**, Circadian rhythm in EAG responses to benzaldehyde (dilution $1:10^3$) in wild-type and *tim*⁰¹ flies on day 2 of DD. Each point represents mean response in 18 flies. Overall effects of time of day ($F_{T5,204} = 20.20$), genotype ($F_{G1,204} = 174.50$) and their interaction ($F_{GT5,204} = 10.92$) are statistically significant ($P < 0.0001$) by two-way ANOVA. Circadian rhythms in EAG responses are robust and statistically significant in wild-type flies, but are abolished in *tim*⁰¹ flies.

effect of genotype was apparent when the entire (day 4) LL data set was considered. Mean responses in wild-type flies were 10–15% greater than those of *per*⁰¹ or *tim*⁰¹ flies. The physiological significance of the slight reduction in mean EAG amplitude in *per*⁰¹ and *tim*⁰¹ flies is unknown, but it does not alter conclusions regarding the circadian regulation of olfactory responses.

Previous studies have demonstrated autonomous rhythms in *per* expression in isolated *Drosophila* antennae, as well as in other peripheral tissues³. In antennae, *Per* protein is expressed in sensory neurons located at the base of the chemosensory sensilla. To date, there have been no reports that peripheral circadian oscillators are essential for any physiological or behavioural rhythm in *Drosophila*. Therefore, in order to determine whether circadian rhythms in olfactory responses require peripheral oscillators, EAG recordings were made from the *per* 7.2:2 strain of transgenic flies in DD. These flies carry a truncated *per* gene in which the promoter and a portion of the first (noncoding) exon are deleted. In this strain, rhythmic *Per* protein expression occurs solely in a subset of the lateral neurons of

the anterior margin of the optic lobe, the site of the pacemaker that controls locomotor behaviour. Importantly, *Per* protein is not expressed in peripheral tissues of *per* 7.2:2 flies, or in central sites outside those lateral neurons essential for locomotor rhythms⁴. We have confirmed that these flies are behaviourally rhythmic (data not shown) as described elsewhere⁴. Although these flies showed robust EAG responses to ethyl acetate in DD, circadian rhythms in EAG responses to ethyl acetate were abolished (Fig. 4). The effects of time of day, genotype and their interaction were statistically significant ($P < 0.0001$), and *post hoc* analyses indicate no significant changes in EAG as a function of time of day in *per* 7.2:2 flies. These results suggest that peripheral oscillators are necessary for circadian rhythms in olfactory responses, and indicate that a complement of central oscillators sufficient to drive locomotor rhythms is insufficient to sustain olfactory rhythms. A different situation may pertain in mammals. For example, circadian rhythms in the expression of the rat *per* homologue (*rPer2*) messenger RNA in peripheral tissues require an intact suprachiasmatic nucleus, and are therefore not autonomous¹¹.

Our experiments indicate that adult wild-type *Drosophila melanogaster* exhibit robust circadian rhythms in electrophysiological responses to two chemically and behaviourally distinct classes of olfactory stimuli. These rhythms are abolished in three *Drosophila* strains that have altered expression of circadian clock genes. Olfactory responses in wild-type flies peaked during the middle of the night, and returned to basal levels just before the end of the night. In contrast, locomotor activity in *Drosophila* is maximal at the time of lights-on and lights-off in LD cycles; the later maximum is the principal peak in DD¹². Surprisingly little is known about the normal behavioural ecology of *Drosophila* outside the laboratory. However, it is possible that olfactory rhythms facilitate the detection of predators, or contribute to opportunistic feeding, at times of day when the animals are relatively inactive. It is also worth noting that complex reproductive behaviour, such as courtship, requires intact olfactory systems¹³, and that olfactory cues are powerful stimuli for associative and non-associative learning, especially in *Drosophila*^{14,15}. The fact that physiological responses to olfactory stimuli are under circadian control suggests that time of day needs to be carefully considered in the design and interpretation of experiments on olfactory learning. □

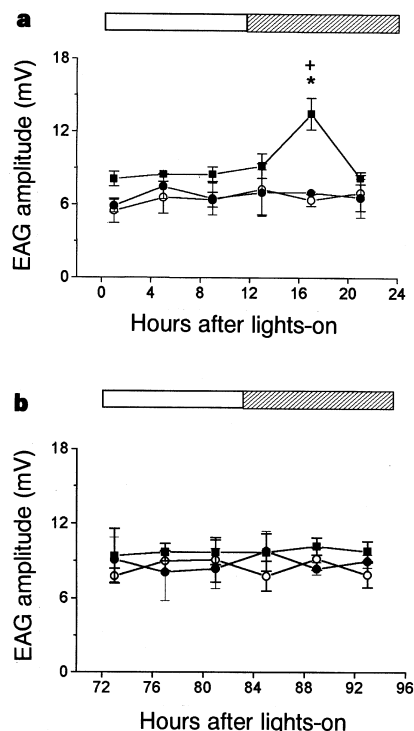


Figure 3 Mean olfactory responses at different times of day in *Drosophila* free-running in constant light (LL). White portion of horizontal bars above graphs indicate day and hatched portions of bars indicate subjective night. Ordinates represent number of hours after lights-on. **a**, Circadian rhythm in olfactory responses of wild-type flies (filled squares) on day 1 of LL. Responses are significantly elevated during the middle of the night (ZT17) compared with other times of day in wild-type, but not *per*⁰¹ (open circles) or *tim*⁰¹ (filled circles) flies. Each point denotes mean EAG response of 24 flies to ethyl acetate (dilution 1:10⁴). Overall effects of time of day ($F_{T5,414} = 9.00$), genotype ($F_{G,2414} = 71.04$) and their interaction ($F_{TG10,414} = 9.23$) are statistically significant ($P < 0.0001$) by two-way ANOVA. Asterisk indicates significant increase ($P < 0.0001$) in EAG responses at ZT17 compared with responses at all other times of day in wild-type flies. Crosses indicate significant increase ($P < 0.05$) in EAG responses in wild-type flies compared with *per*⁰¹ and *tim*⁰¹ flies at the same times of day. **b**, Circadian rhythm in olfactory responses is fully damped by day 4 of LL in all three strains of flies. Points denote mean EAG response in 12 flies. The overall effect of time of day is not statistically significant ($F_{T5,414} = 0.45$; $P > 0.05$), but the overall effect of genotype is statistically significant ($F_{G,2,414} = 12.78$; $P < 0.001$). Overall mean responses in wild-type flies are greater than those of *per*⁰¹ or *tim*⁰¹ flies when data from all time points are pooled.

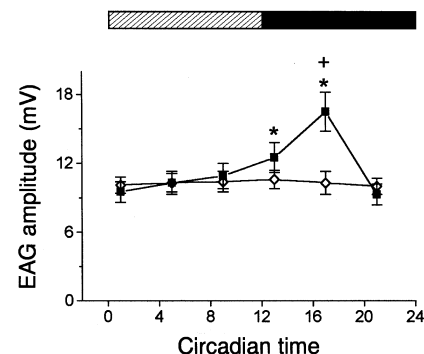


Figure 4 Olfactory responses in wild-type (squares) and *per* 7.2:2 transgenic *Drosophila* (diamonds) during day 2 of DD. Each point represents mean EAG response to ethyl acetate (dilution 1:10⁴) in 24 flies. Overall effects of time of day ($F_{T5,414} = 46.52$), genotype ($F_{G,5,277} = 54.61$) and their interaction ($F_{TG5,277} = 40.59$) are statistically significant ($P < 0.0001$) by two-way ANOVA. *Post hoc* analyses indicate no significant differences as a function of the time of day in the *per* 7.2:2 transgenic line, suggesting that peripheral circadian oscillators are necessary for the circadian rhythm. Asterisks indicate significant ($P < 0.05$) increase in EAG responses at CT13 and CT17 compared with responses at all other times of day in wild-type flies. Cross indicates significant ($P < 0.001$) increase in EAG response in wild-type flies compared to *per* 7.2:2 flies at CT17.

Methods

Adult female *Drosophila* (2–4 days post-eclosion) were entrained to 12-h light:12-h dark (LD) cycles for 3 days at 25°C. Lights-on occurred at ZT0 and lights-off at ZT12. EAG recordings were made on day 4 of LD, on days 1 and 4 of LL and on day 2 of DD. In different repetitions of these experiments, the entraining light–dark cycles had different phase relationships to ambient light–dark cycles so as to eliminate possible rhythmic contributions from uncontrolled environmental variables. Wild-type *Drosophila* were taken from a Canton-S strain. The *per* 7.2:2 transgene was carried in a *per*⁰¹ genetic background (*per*⁰¹; P(7.2:2)ry). For electrophysiology, *Drosophila* were immobilized 1 h before recording. Odorant-evoked changes in transepithelial potential (EAG responses) were monitored by an indifferent electrode inserted into the lumen of the antenna and a recording electrode pressed against the extracuticular surface on the medial–anterior face of the third antennal segment. Electrodes were filled with 0.17 M NaCl (resistance of 2–10 MΩ). Electrode placement in the dark was made possible by an IR light source (620 nm cutoff) and a microscope equipped with an IR-sensitive video camera. The *Drosophila* circadian system is insensitive to light above 600 nm (ref. 16). The preparation was shielded from light sources in the room. EAG responses were recorded with a high-input impedance electrometer and filtered at 2 kHz with an 8-pole Bessel filter. Temperature was 22–24°C at the time of recording and did not vary systematically as a function of the time of day. Ethyl acetate (Sigma) or benzaldehyde (Fluka) were diluted in paraffin and applied to animals by a system of manually operated valves that directed a constant flow of air (250 ml min⁻¹) through various reservoirs, each containing a different odorant dilution (ranging from 1:10¹⁰ to 1:10⁴ in the case of ethyl acetate and 1:10⁹ to 1:10³ in the case of benzaldehyde), and from there to the fly. Separate odorant delivery systems were used for benzaldehyde or ethyl acetate. Air was constantly flowing from a port 5 mm from the fly during EAG recording. Different odorant dilutions were applied in random order once every 30 s. Analyses of the effects of time of day, genotype and their interactions were analysed by two-way ANOVA. *Post hoc* analyses of individual treatment groups were then performed using Scheffé's test ($\alpha = 0.05$). Statistical analyses were carried out with STATISTICA software (StatSoft; Tulsa). For analysis, each treatment group consisted of 12–24 flies pooled from several replications of experiments performed on smaller groups.

Received 26 April; accepted 26 May 1999.

- Hardin, P. E. Activating inhibitors and inhibiting activators: A day in the life of a fly. *Curr. Opin. Neurobiol.* **8**, 642–647 (1998).
- Dunlap, J. C. Molecular basis for circadian clocks. *Cell* **96**, 271–290 (1999).
- Plautz, J. D., Kaneko, M., Hall, J. C. & Kay, S. A. Independent photoreceptive circadian clocks throughout *Drosophila*. *Science* **278**, 1632–1635 (1997).
- Frisch, B., Hardin, P. E., Hamblen-Coyle, M. J., Rosbash, M. & Hall, J. C. A promoterless *period* gene mediates behavioral rhythmicity and cyclical *per* expression in a restricted subset of the *Drosophila* nervous system. *Neuron* **12**, 555–570 (1994).
- Alcorta, E. Characterization of the electroantennogram in *Drosophila melanogaster* and its use for identifying olfactory capture and transduction mutants. *J. Neurophysiol.* **65**, 702–714 (1991).
- Ayer, R. K. & Carlson, J. R. Olfactory physiology in *Drosophila* antenna and maxillary palp: *Acj6* distinguishes two classes of odorant pathways. *J. Neurobiol.* **23**, 965–982 (1992).
- Dubin, A. E., Heald, N. L., Cleveland, B., Carlson, J. R. & Harris, G. L. *Scutoid* mutation of *Drosophila melanogaster* specifically decreases olfactory responses to short-chain acetate esters and ketones. *J. Neurobiol.* **28**, 214–233 (1995).
- Anholt, R. R., Lyman, R. F. & Mackay, T. F. Effects of single P-element insertions on olfactory behavior in *Drosophila melanogaster*. *Genetics* **143**, 293–301 (1996).
- Qiu, J. & Hardin, P. E. *per* mRNA cycling is locked to lights-off under photoperiodic conditions that support circadian feedback loop function. *Mol. Cell. Biol.* **6**, 4182–4188 (1996).
- Marrus, S. B., Zheng, H. & Rosbash, M. Effect of constant light and circadian entrainment of *per*^S flies: Evidence for light-mediated delay of the negative feedback loop in *Drosophila*. *EMBO J.* **15**, 6877–6886 (1996).
- Sakamoto, K. *et al.* Multitissue circadian expression of rat *period* homolog (*rPer2*) mRNA is governed by the mammalian circadian clock, the suprachiasmatic nucleus in the brain. *J. Biol. Chem.* **273**, 27039–27042 (1998).
- Konopka, R. J. & Benzer, S. Clock mutants of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **68**, 2112–2116 (1971).
- Hall, J. C. The mating of a fly. *Science* **264**, 1702–1714 (1994).
- Tully, T. & Quinn, W. G. Classical conditioning and retention in normal and mutant *Drosophila melanogaster*. *J. Comp. Physiol. A* **157**, 263–277 (1985).
- Davis, R. L. Physiology and biochemistry of *Drosophila* learning mutants. *Physiol. Rev.* **76**, 299–317 (1996).
- Suri, V., Qian, Z., Hall, J. C. & Rosbash, M. Evidence that the TIM light response is relevant to light-induced phase shifts in *Drosophila melanogaster*. *Neuron* **21**, 225–234 (1998).

Acknowledgements. This work was supported by grants from the NIH to S.E.D. and P.E.H. B.K. received support from the BRAIN program. We thank L. Dryer and N. Glossop for discussions and comments on the manuscript. S.E.D. and P.E.H. are NHTS.

Correspondence and requests for materials should be addressed to S.E.D. (e-mail: SDryer@uh.edu).

Interleukin-4-dependent production of PPAR- γ ligands in macrophages by 12/15-lipoxygenase

Jannet T. Huang^{*†}, John S. Welch^{*†}, Mercedes Ricote^{*†}, Christoph J. Binder^{*}, Timothy M. Willson[‡], Carolyn Kelly[§], Joseph L. Witztum^{*}, Colin D. Funk^{||}, Douglas Conrad[†] & Christopher K. Glass^{*#}

^{*} Division of Endocrinology and Metabolism, [§] Division of Nephrology, [¶] Division of Pulmonary and Critical Care Medicine, [#] Division of Cellular and Molecular Medicine, Department and School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0651, USA [‡] Department of Medicinal Chemistry, Glaxo Wellcome Research and Development, P.O. Box 13398 Research Triangle Park, North Carolina 27709, USA

^{||} Center for Experimental Therapeutics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[†] These authors contributed equally to this work

The peroxisome proliferator-activated receptor- γ (PPAR- γ) is a ligand-dependent nuclear receptor that has been implicated in the modulation of critical aspects of development and homeostasis, including adipocyte differentiation¹, glucose metabolism^{2,3} and macrophage development and function^{4–6}. PPAR- γ is activated by a range of synthetic and naturally occurring substances, including antidiabetic thiazolidinediones^{2,3}, polyunsaturated fatty acids⁷, 15-deoxy- $\Delta^{12,14}$ prostaglandin J₂ (refs 8, 9) and components of oxidized low-density lipoprotein, such as 13-hydroxyoctadecadienoic acid (13-HODE) and 15-hydroxyicosatetraenoic acid (15-HETE)¹⁰. However, the identities of endogenous ligands for PPAR- γ and their means of production *in vivo* have not been established. In monocytes and macrophages, 13-HODE and 15-HETE can be generated from linoleic and arachidonic acids, respectively, by a 12/15-lipoxygenase that is upregulated by the T_H2-derived cytokine interleukin-4 (ref. 11). Here we show that interleukin-4 also induces the expression of PPAR- γ and provide evidence that the coordinate induction of PPAR- γ and 12/15-lipoxygenase mediates interleukin-4-dependent transcription of the CD36 gene in macrophages. These findings reveal a physiological role of 12/15-lipoxygenase in the generation of endogenous ligands for PPAR- γ , and suggest a paradigm for the regulation of nuclear receptor function by cytokines.

PPAR- γ is highly expressed in macrophages obtained from thioglycollate-elicited peritonitis but not in resident peritoneal macrophages, indicating that humoral factors produced during inflammation stimulate PPAR- γ expression (Fig. 1a). We therefore evaluated cytokines and other factors known to regulate macrophage differentiation and function for their potential effects on PPAR- γ expression. Among the factors tested, interleukin-4 (IL-4) strongly induced PPAR- γ messenger RNA and protein in resident peritoneal macrophages (Fig. 1a, c). Granulocyte–macrophage colony-stimulating factor (GM-CSF) and macrophage colony-stimulating factor (M-CSF) also upregulated PPAR- γ , as previously described¹², but did not act synergistically with IL-4 (Fig. 1a). IL-4 also significantly induced PPAR- γ expression in human peripheral blood monocytes (Fig. 1b). Other inflammatory mediators, including IL-1 β , IL-10, interferon- γ , tumour-necrosis factor- α and lipopolysaccharide either had no effect on murine PPAR- γ expression or were inhibitory (data not shown). PPAR- γ is controlled by three different promoters that direct expression of PPAR- γ 1, PPAR- γ 2 and PPAR- γ 3 mRNAs^{13,14}, which may allow tissue-specific differential regulation. IL-4 exclusively induced the expression of the

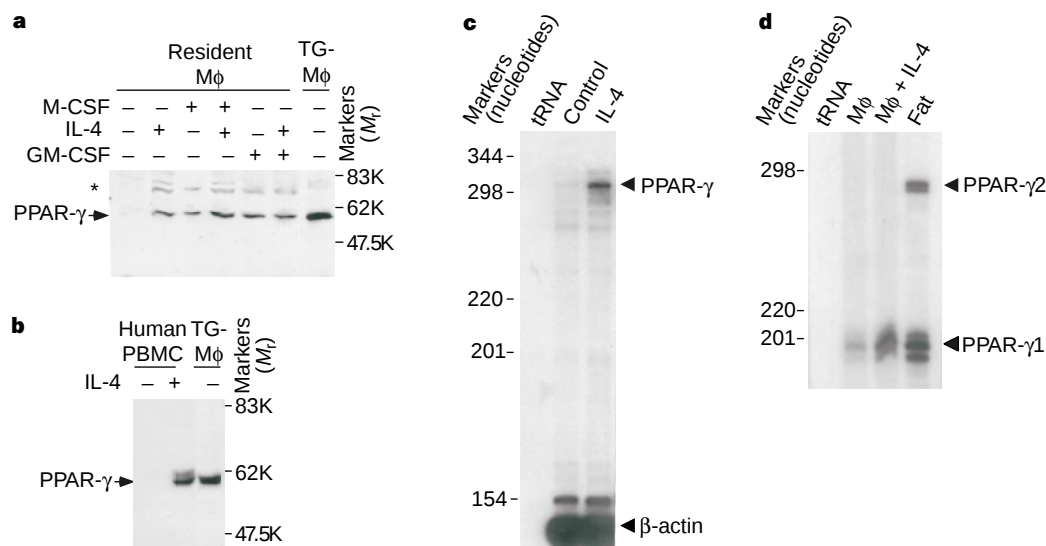


Figure 1 Regulation of PPAR-γ expression in macrophages by IL-4. **a**, Regulation of PPAR-γ protein expression. TG-Mφ represent thioglycollate-elicited macrophages. The asterisk denotes a nonspecific band. Equivalent protein loading was confirmed by a parallel western blot using an anti-β-actin antibody. **b**, IL-4-induced PPAR-γ protein expression in human peripheral blood monocytes (Human PBMC). **c**, Regulation of PPAR-γ mRNA expression. The specific 293-nucleotide PPAR-γ protection product is indicated by an arrowhead. Equivalent loading of RNA was demonstrated by an antisense probe hybridizing to β-Actin

with a protected fragment of 143 nucleotides. **d**, IL-4 induces expression of PPAR-γ1. 194 and 284 nucleotide protected fragments corresponding to PPAR-γ1 and PPAR-γ2 were obtained from fat-derived RNA, but only the 194 nucleotide fragment corresponding to PPAR-γ1 was induced in IL-4-treated macrophages. Equivalent loading of RNA was confirmed by a parallel RNase protection assay using an antisense probe recognizing β-actin. RNase protection assays using a probe that can distinguish PPAR-γ3 and PPAR-γ1 transcripts indicated that PPAR-γ3 is not expressed in IL-4-induced macrophages (data not shown).

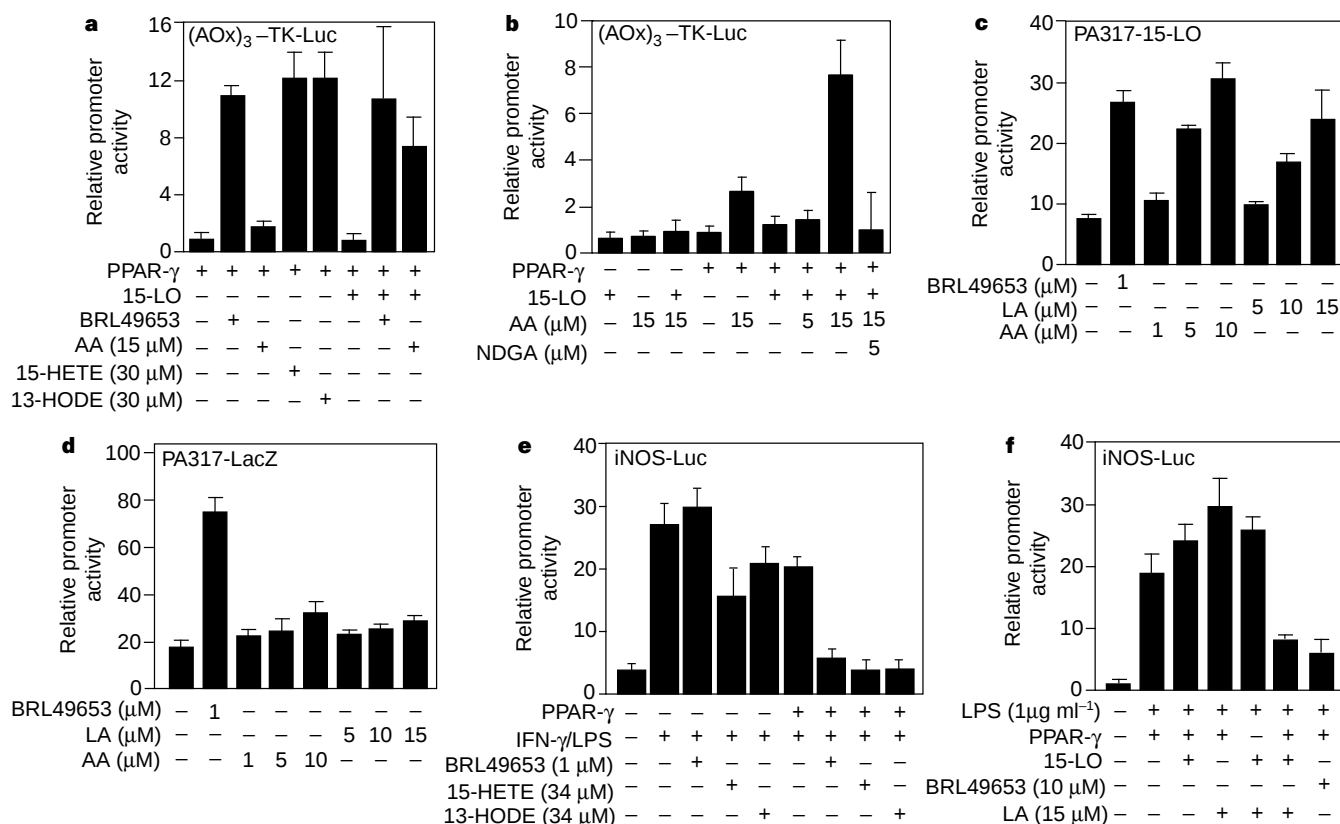


Figure 2 15-lipoxygenase products of arachidonic acid and linoleic acid metabolism potentiate transcriptional activation of PPAR-γ. **a**, **b**, Expression of 15-LO enhanced the ability of arachidonic acid to activate a PPAR-γ-dependent reporter gene (AOX)₃-TK-Luc in RAW 264.7 cells. This effect required PPAR-γ and was inhibited by NDGA. **c**, **d**, Stable expression of 15-LO potentiates transcriptional activation of PPAR-γ by arachidonic acid (AA) and linoleic acid (LA) in PA317

cells. PA317-15-LO and PA317-LacZ are cell lines¹⁸ that stably express 15-LO or lacZ, respectively. **e**, BRL49653, 13-HETE and 15-HODE inhibited induction of the iNOS promoter in a PPAR-γ-dependent manner in RAW264.7 cells. **f**, Linoleic acid inhibited the iNOS promoter in the presence of co-expressed 15-LO and PPAR-γ. Error bars represent s.d.

PPAR- γ 1 isoform in murine peritoneal macrophages, as shown using antisense mRNA probes capable of distinguishing among the PPAR- γ 1, γ 2 and γ 3 transcripts (Fig. 1d, and data not shown).

In human monocytes, IL-4 induces the expression of 15-lipoxygenase (15-LO), a lipid-peroxidating enzyme that can oxygenate free polyunsaturated fatty acids and phospholipids in biomembranes¹⁵. 15-LO has been implicated in a number of cellular processes, including degradation of intracellular organelles¹⁶ and oxidation of low-density lipoprotein (LDL)^{17,18}, but its physiological role remains poorly understood. A leucocyte 12/15-lipoxygenase (12/15-LO) found in murine macrophages is highly homologous with human 15-LO at the levels of protein, enzymatic and molecular biological characteristics; it is therefore thought to be the species equivalent of human 15-LO^{19,20}. In mammalian cells, linoleic acid and arachidonic acid are the major polyunsaturated fatty acid substrates for 12/15-LO²⁰. Linoleic acid is oxygenated by IL-4 treated human monocytes to 13-HODE²⁰, whereas 12-HETE and 15-HETE are the major products of the 12/15-LO pathway in murine macrophages when arachidonic acid is added as the substrate²¹. 13-HODE activates PPAR- γ when added to cells at high concentrations (for example, 30–50 μ M)¹⁰. The finding that IL-4 induced PPAR- γ expression in macrophages indicated a possible link between the expression of 12/15-LO and the production of activating ligands for PPAR- γ . To investigate this possibility, we co-transfected PPAR- γ and 15-LO expression vectors into the murine macrophage RAW 264.7 cell line. In the absence of co-transfected 15-LO, addition of arachidonic acid resulted in only a small increase in the transcription of a PPAR- γ -dependent reporter gene ((AOx)₃-TK-Luc), consistent with a low level of endogenous 12/15-LO activity in these cells²². However, co-expression of 15-LO significantly enhanced the ability of arachidonic acid to activate this reporter gene (Fig. 2a). This effect required co-expression of PPAR- γ , was inhibited by nordihydroguaiaretic acid (NDGA) (Fig. 2b) and was dependent on the amount of co-transfected 15-LO expression vector (data not shown). 13-HODE and 15-HETE directly activated reporter gene transcription, consistent with previous findings¹⁰ (Fig. 2a). PPAR- γ -dependent transcription was also significantly enhanced by arachidonic acid and linoleic acid in murine PA317 fibroblasts that stably express human 15-LO protein and show high enzyme activity, but not in control cells that stably express β -galactosidase and have no detectable 15-LO activity¹⁸ (Fig. 2c, d). These findings indicate that 15-LO can effectively catalyse the conversion of arachidonic acid and linoleic acid to activating ligands of PPAR- γ .

Products of the 12/15-LO pathway also stimulate transrepression activity of PPAR- γ . PPAR- γ ligands inhibit the induction of inducible nitric oxide synthase (iNOS) by lipopolysaccharide and interferon- γ (IFN- γ) (ref. 5). IL-4 can inhibit induction of iNOS in primary mouse macrophages²³, so IL-4 may inhibit iNOS by inducing PPAR- γ and its ligands. Consistent with this hypothesis, 15-HETE and 13-HODE inhibited induction of the iNOS promoter in a PPAR- γ -dependent manner in RAW 264.7 cells, and linoleic acid inhibited the iNOS promoter in the presence of co-expressed 15-LO and PPAR- γ (Fig. 2e, f).

To examine the possibility that IL-4 affects macrophage gene expression by coordinate induction of PPAR- γ and 12/15-LO, we investigated the potential roles of these proteins in mediating IL-4 stimulation of CD36 (ref. 24). CD36 is a multifunctional protein that is expressed in diverse cell types, including platelets, adipocytes and macrophages²⁵. In macrophages, CD36 may function as a scavenger receptor for oxidized LDL²⁶ and apoptotic cells²⁷, and its expression in these cells is positively regulated by PPAR- γ ^{4,10}. Consistent with these findings, IL-4 and various PPAR- γ ligands including 13-HODE, 15-HETE, troglitazone, BRL49653 and 15-deoxy- $\Delta^{12,14}$ prostaglandin J₂ (15dPGJ₂) each stimulated expression of CD36 in thioglycollate-elicited peritoneal macrophages (Fig. 3a, b). Activation of CD36 by IL-4 was blocked by pretreatment with

the specific 12/15-LO inhibitor PD146176 (Fig. 3c) as well as by NDGA and by baicalein (data not shown). In addition, an irreversible antagonist of PPAR- γ , GW9662, also inhibited activation of CD36 by IL-4 (Fig. 3d). The antagonistic activity of GW9662 was shown by its ability to block BRL49653 activation of PPAR- γ on the (AOx)₃-TK-Luc promoter in a dose-dependent manner (Fig. 3e). These inhibitors had no effect on overall protein expression or the expression of β -actin, indicating specific effects on gene expression at the concentrations used. Furthermore, the inhibitory effects of PD146176 could be rescued by treatment with troglitazone (Fig. 3c).

To confirm a role of 12/15-LO in IL-4-dependent induction of CD36, we examined CD36 expression in thioglycollate-elicited peritoneal macrophages obtained from 12/15-LO-deficient mice. Although these mice lack 12/15-LO mRNA, protein and enzyme activity, they are fertile, grow normally and do not exhibit overt haematopoietic abnormalities^{19,28}. IL-4 induction of CD36 was significantly reduced in macrophages from 12/15-LO knockout mice, compared to wild-type mice (Fig. 4a). In most experiments,

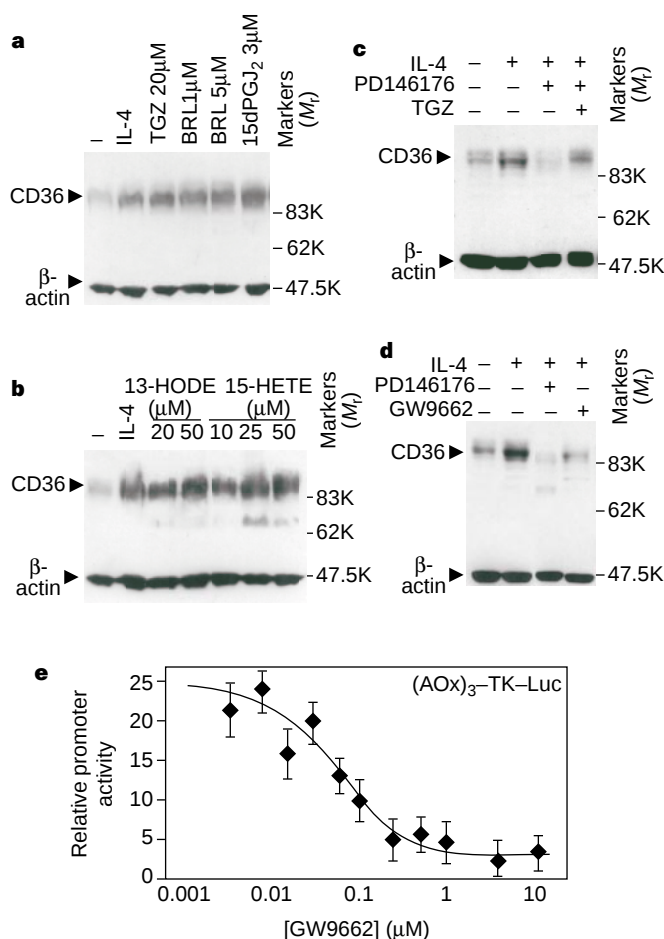


Figure 3 Inhibitors of 15-LO and PPAR- γ block IL-4-dependent expression of CD36 in macrophages. **a**, Expression of CD36 (detected by western blot) is induced in peritoneal macrophages by IL-4, troglitazone (TGZ), BRL49653 and 15dPGJ₂. **b**, Expression of CD36 is increased in thioglycollate-elicited peritoneal macrophages by 13-HODE and 15-HETE. **c**, CD36 protein levels were upregulated by IL-4 and this effect was blocked by pretreatment with the 15-LO inhibitor PD146176 (10 μ M). The reduced CD36 expression by PD146176 (1 μ M) in IL-4-treated macrophages was rescued by the PPAR- γ ligand troglitazone (15 μ M). **d**, IL-4-dependent induction of CD36 was blunted by the PPAR- γ antagonist GW9662 (1 μ M). **e**, GW9662 antagonized PPAR- γ activation of the transfected (AOx)₃-TK-Luc reporter gene by BRL49653 in a dose-dependent manner in RAW264.7 cells.

basal expression of CD36 was also significantly reduced in 12/15-LO knockout macrophages, consistent with lower levels of endogenous PPAR- γ ligands as compared to wild-type macrophages. This interpretation is further supported by the finding that treatment of 12/15-LO^{-/-} macrophages with troglitazone rescued CD36 expression, indicating that there may be a selective defect in the production of endogenous PPAR- γ ligands (Fig. 4a). As IL-4 retained a partial ability to stimulate CD36 expression in 12/15-LO^{-/-} macrophages, 12/15-LO-independent mechanisms also appear to contribute to IL-4 activation of CD36. We used the cyclooxygenase inhibitor aspirin and the 5-LO inhibitor caffeic acid to investigate the involvement of these enzymes in the IL-4 induction of CD36. Neither of these compounds reduced CD36 expression in either

wild-type or 12/15-LO^{-/-} peritoneal macrophages (Fig. 4b). Thus, although 15dPGJ₂ is a naturally occurring substance that can activate PPAR- γ , prostaglandin J₂ metabolites do not appear to contribute to expression of CD36 under these conditions.

The PPAR- γ ligand-binding pocket is two to three times larger than those of previously characterized steroid hormone receptors, presumably accounting for its ability to accommodate a broad spectrum of structurally diverse molecules²⁹. Although many naturally occurring fatty acids and their metabolites can activate PPAR- γ , they bind with relatively low affinities and must be added to cells at high concentrations to stimulate transcription. It has therefore been difficult to establish the physiological relevance of any of these substances as regulators of PPAR- γ *in vivo*. The finding that inhibition of 12/15-LO or disruption of the 12/15-LO gene significantly reduced CD36 expression supports an important role for structurally related 12/15-LO products, including 15-HETE and 13-HODE, as endogenous regulators of PPAR- γ in the macrophage. Coordinate induction of both PPAR- γ and 12/15-LO synthesis by IL-4 establishes a novel paradigm for cytokine regulation of nuclear receptor function that may provide a mechanism for tissue-specific activation of transcription (Fig. 4c). Finally, our results indicate an alternative strategy for the identification of endogenous ligands for orphan nuclear receptors based on pharmacological and genetic manipulation of specific metabolic pathways. □

Methods

Cell culture. Primary resident and thioglycollate-elicited peritoneal macrophages were isolated and cultured as described⁵. RAW 264.7 and PA317 cells were grown as described^{12,18}. We obtained human peripheral blood monocytes from healthy donors as described¹¹ and cultured them in RPMI with 10% autologous serum. Recombinant cytokines from Endogen were used at the following concentrations: human macrophage colony-stimulating factor, 20 ng ml⁻¹; murine granulocyte-macrophage colony-stimulating factor, 5 ng ml⁻¹; murine IL-1 β , 10 ng ml⁻¹; murine IL-4, 10 ng ml⁻¹; human IL-4, 10 ng ml⁻¹; murine IL-10, 10 ng ml⁻¹; murine tumour-necrosis factor- α , 50 ng ml⁻¹; murine interferon- γ , 100 U ml⁻¹ and lipopolysaccharide, 5 μ g ml⁻¹. Troglitazone, BRL49653 and GW9662 were synthesized at Glaxo Wellcome. Arachidonic acid, linoleic acid, 15-HETE and 13-HODE were from Cayman Chemical; baicalein from Biomol; caffeic acid and NDGA from Aldrich; and aspirin from Sigma. PD146176 was a gift from J. Cornicelli of Parke-Davis.

Transient transfection. We carried out transient transfections as described using Lipofectamine⁵ to transfect the RAW 264.7 and PA317 cell lines. The iNOS promoter-luciferase construct, PPAR- γ -dependent reporter construct (AOx)₃-TK-luciferase and PPAR- γ expression vector pcMX-PPAR- γ have been described previously⁵. The *EcoRV/NotI* fragment of the 15-LO complementary DNA was cloned into pcDNA3 (Invitrogen) to generate a 15-LO expression vector³⁰. Transfections were performed at 2 $\times$ 10⁵ cells per well in six-well plates. The cells were allowed to rest overnight in medium containing 0.5% FBS, followed by treatment with the indicated compounds for 24 h. We assayed luciferase activity in RAW 264.7 cells as described⁵ or using the Dual-Luciferase reporter assay system (Promega) according to the manufacturer's directions.

Western blot analysis. Western blot analysis was done using standard procedures¹². Resident peritoneal macrophages and thioglycollate-elicited macrophages were treated with the various compounds indicated in each experiment for 48 h. Human peripheral blood monocytes were cultured with or without IL-4 for 96 h. Total protein was extracted in SDS sample buffer, resolved by SDS-PAGE and transferred to nitrocellulose filters. To detect the 55K PPAR- γ protein, incubation with a guinea-pig antiserum against full-length murine PPAR- γ fused to glutathione-S-transferase was performed at 1:500 dilution overnight at 4 °C. Alternatively, a murine monoclonal antibody against PPAR- γ (Santa Cruz, E8) was used at 1:200 for 1 h at room temperature. We used a guinea-pig antiserum against the recombinant extracellular domain of murine CD36 at 1:1000 dilution to detect of the 88K CD36 protein. Secondary rabbit anti-guinea pig (Dako) and goat anti-mouse (Dako) antibodies were diluted at 1:10,000 and 1:2500, respectively. For detection we used chemiluminescence (Pierce). Protein levels were normalized using a mouse monoclonal antibody against β -actin (Sigma).

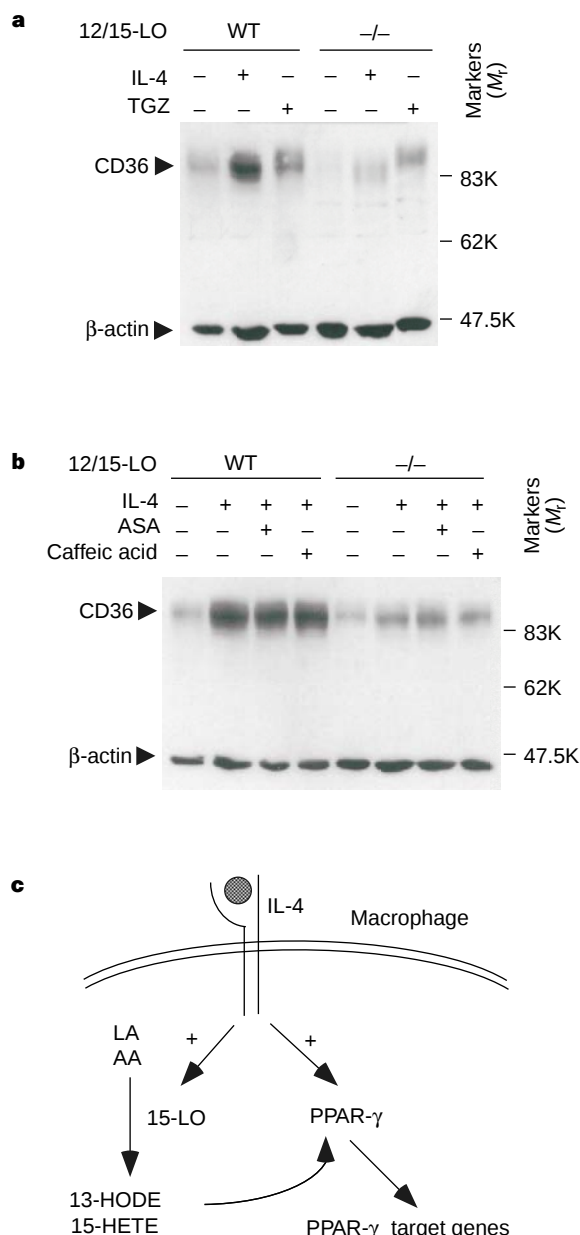


Figure 4 Induction of CD36 expression by IL-4 is defective in 12/15-LO^{-/-} macrophages. **a**, Basal and IL-4-dependent expression of CD36 were diminished in 12/15-LO^{-/-} mice, with significant expression restored by treatment with troglitazone (15 μ M). **b**, CD36 expression is not influenced by aspirin (ASA) (200 μ M) or caffeic acid (20 μ M) in both wild-type and 12/15-LO^{-/-} macrophages. **c**, Model for IL-4 regulation of PPAR- γ function: IL-4 induces target gene expression by simultaneously enhancing PPAR- γ levels and the production of activating ligands through 15-LO oxidation of linoleic and arachidonic acids.

RNase protection assay. For the specific analysis of the PPAR- γ 3 relative to PPAR- γ 1 mRNA, we constructed a template for probe synthesis using an RT-PCR product derived from mouse adipose tissue RNA with oligonucleotides that bind sense at exon A1 and antisense at exon 2. The amplified fragment was inserted blunt into pCR2.1 (TA cloning, Invitrogen). The resulting plasmid pCR2.1-PPAR- γ 3-RPA, which contains part of exon A1, the full-length exon A2, exon 1 and part of exon 2, was used as a template for probe synthesis. To analyse PPAR- γ 2 and PPAR- γ 1, the coding region from +1 to +284 was obtained as an RT-PCR product from mouse adipose tissue RNA and cloned into pCR2.1, then subcloned into pBluescript KS+ (Stratagene). We used the resulting plasmid as a template for synthesis of the antisense RNA probe, allowing detection of PPAR- γ 2 and PPAR- γ 1 mRNA. The common PPAR- γ probe corresponding to nucleotides 800–1093 of the murine cDNA was subcloned into pBluescript SK+ (Stratagene). RNase protection experiments were done as described¹². Resident peritoneal macrophages were treated in the presence or absence of IL-4 for 18 h. We used a β -actin antisense probe to verify equivalent amounts of total mRNA.

Received 28 April; accepted 20 May 1999.

1. Tontonoz, P., Hu, E. & Spiegelman, B. M. Stimulation of adipogenesis in fibroblasts by PPAR γ 2, a lipid-activated transcription factor. *Cell* **79**, 1147–1156 (1994).
2. Lehmann, J. M. *et al.* An antidiabetic thiazolidinedione is a high affinity ligand for peroxisome proliferator-activated receptor- γ (PPAR γ). *J. Biol. Chem.* **270**, 12953–12956 (1995).
3. Willson, T. M. *et al.* The structure–activity relationship between peroxisome proliferator-activated receptor γ agonism and the antihyperglycemic activity of thiazolidinediones. *J. Med. Chem.* **39**, 665–668 (1996).
4. Tontonoz, P., Nagy, L., Alvarez, J. G. A., Thomazy, V. A. & Evans, R. M. PPAR γ promotes monocyte/macrophage differentiation and uptake of oxidized LDL. *Cell* **93**, 241–252 (1998).
5. Ricote, M., Li, A. C., Willson, T. M., Kelly, C. J. & Glass, C. K. The peroxisome proliferator-activated receptor- γ is a negative regulator of macrophage activation. *Nature* **391**, 79–82 (1998).
6. Jiang, C., Ting, A. T. & Seed, B. PPAR- γ agonists inhibit production of monocyte inflammatory cytokines. *Nature* **391**, 82–86 (1998).
7. Kliewer, S. A. *et al.* Fatty acids and eicosanoids regulate gene expression through direct interactions with peroxisome proliferator-activated receptors α and γ . *Proc. Natl Acad. Sci. USA* **94**, 4318–4323 (1997).
8. Kliewer, S. A. *et al.* A prostaglandin J_2 metabolite binds peroxisome proliferator-activated receptor and promotes adipocyte differentiation. *Cell* **83**, 813–819 (1995).
9. Forman, B. M. *et al.* 15-Deoxy- $\Delta^{12,14}$ -prostaglandin J_2 is a ligand for the adipocyte determination factor PPAR γ . *Cell* **83**, 803–812 (1995).
10. Nagy, L., Tontonoz, P., Alvarez, J. G. A., Chen, H. & Evans, R. M. Oxidized LDL regulates macrophage gene expression through ligand activation of PPAR- γ . *Cell* **93**, 229–240 (1998).
11. Conrad, D. J., Kuhn, H., Mulkins, M., Highland, E. & Sigal, E. Specific inflammatory cytokines regulate the expression of human monocyte 15-lipoxygenase. *Proc. Natl Acad. Sci. USA* **89**, 217–221 (1992).
12. Ricote, M. *et al.* Expression of the peroxisome proliferator-activated receptor γ (PPAR γ) in human atherosclerosis and regulation in macrophages by colony stimulating factors and oxidized low density lipoprotein. *Proc. Natl Acad. Sci. USA* **95**, 7614–7619 (1998).
13. Zhu, Y. *et al.* Structural organization of mouse peroxisome proliferator-activated receptor γ (mPPAR γ) gene: alternative promoter use and different splicing yield two mPPAR γ isoforms. *Proc. Natl Acad. Sci. USA* **92**, 7921–7925 (1995).
14. Fajas, L., Fruchart, J. C. & Auwerx, J. PPAR γ 3 mRNA: a distinct PPAR γ mRNA subtype transcribed from an independent promoter. *FEBS Lett.* **438**, 55–60 (1998).
15. Yamamoto, S. Mammalian lipoxygenases: molecular structure and functions. *Biochem. Biophys. Res. Commun.* **1128**, 117–131 (1992).
16. van Leyen, K., Duvoisin, R. M., Engelhardt, H. & Wiedmann, M. A function for lipoxygenase in programmed organelle degradation. *Nature* **395**, 392–395 (1998).
17. Rankin, S. M., Parthasarathy, S. & Steinberg, D. Evidence for a dominant role of lipoxygenase(s) in the oxidation of LDL by mouse peritoneal macrophages. *J. Lipid Res.* **32**, 449–456 (1991).
18. Benz, D. J. *et al.* Enhanced levels of lipoperoxides in low density lipoproteins incubated with murine fibroblasts expressing high levels of human 15-lipoxygenase. *J. Biol. Chem.* **270**, 5191–5197 (1995).
19. Funk, C. D. The molecular biology of mammalian lipoxygenases and the quest for eicosanoid functions using lipoxygenase-deficient mice. *Biochim. Biophys. Acta* **1304**, 65–84 (1996).
20. Kuhn, H. Biosynthesis, metabolism and biological importance of the primary 15-lipoxygenase metabolites 15-hydro(pero)xy-5Z,8Z,11Z,13E-eicosatetraenoic acid and 13-hydro(pero)xy-9Z,11E-octadecadienoic acid. *Prog. Lipid Res.* **35**, 203–226 (1996).
21. Heydeck, D. *et al.* Interleukin-4 and -13 induce upregulation of the murine macrophage 12/15-lipoxygenase activity: evidence for the involvement of transcription factor STAT6. *Blood* **92**, 2503–2510 (1998).
22. Scheidegger, K. J., Butler, S. & Witztum, J. L. Angiotensin II increases macrophage-mediated modification of low density lipoprotein via a lipoxygenase-dependent pathway. *J. Biol. Chem.* **272**, 21609–21615 (1997).
23. Bogdan, C., Vodovotz, Y., Paik, J., Xie, Q.-W. & Nathan, C. Mechanism of suppression of nitric oxide synthase expression by interleukin-4 in primary mouse macrophages. *J. Leukocyte Biol.* **55**, 227–233 (1994).
24. Yesner, L. M., Huh, H. Y., Pearce, S. F. & Silverstein, R. L. Regulation of monocyte CD36 and thrombospondin-1 expression by soluble mediators. *Arterioscler. Thromb. Vasc. Biol.* **16**, 1019–1025 (1996).
25. Yamada, Y., Doi, T., Hamakubo, T. & Kodama, T. Scavenger receptor family proteins: roles for atherosclerosis, host defence and disorders of the central nervous system. *Cell. Mol. Life Sci.* **54**, 628–640 (1998).
26. Endemann, G. *et al.* CD36 is a receptor for oxidized low density lipoprotein. *J. Biol. Chem.* **268**, 11811–11816 (1993).
27. Savill, J. S., Hogg, N., Ren, Y. & Haslett, C. Thrombospondin cooperates with CD36 and the vitronectin receptor in macrophage recognition of neutrophils undergoing apoptosis. *J. Clin. Invest.* **90**, 1513–1522 (1992).
28. Sun, D. & Funk, C. D. Disruption of 12/15-lipoxygenase expression in peritoneal macrophages. *J. Biol. Chem.* **271**, 24055–24062 (1996).

29. Nolte, R. T. *et al.* Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor- γ . *Nature* **395**, 137–143 (1998).
30. Sigal, E. C., Craik, C. S. & Highland, E. Molecular cloning and primary structure of human 15-lipoxygenase. *Biochem. Biophys. Res. Commun.* **157**, 457–464 (1988).

Acknowledgements. We thank S. Blanchard (Glaxo Wellcome) for communicating results of the characterization of GW9662 before publication; T. Schneiderman for assistance with manuscript preparation; and J. Barnett for assistance in preparation of human peripheral blood monocytes. J.T.H. is supported by an NIH Postdoctoral Fellowship. J.S.W. is supported by an NIH Medical Scientist Training Program Grant to U. C. San Diego. M.R. is supported by a Postdoctoral Fellowship from the American Heart Association. C.J.K. and D.C. are clinical investigators of the Medical Research Service, Department of Veterans Affairs. C.K.G. is an Established Investigator of the American Heart Association. These studies were also supported by NIH grants to C.D.F. and C.K.G. and a Specialized Center of Research Grant in Atherosclerosis to J.L.W. and C.K.G.

Correspondence and requests for materials should be addressed to C.K.G. (e-mail: cglass@ucsd.edu).

Regulation of cell movement is mediated by stretch-activated calcium channels

Juliet Lee^{*†}, Akira Ishihara^{*}, Gerry Oxford[‡], Barry Johnson[§] & Ken Jacobson^{*||}

^{*} Department of Cell Biology and Anatomy, ^{||} Lineberger Comprehensive Cancer Center and [‡] Department of Molecular and Cell Physiology, University of North Carolina, Chapel Hill, North Carolina 27599-7090, USA
[§] Department of Physiology and Neurobiology, University of Connecticut, 75 North Eagleville Road, U-156, Storrs, Connecticut 06269, USA

Intracellular calcium regulates many of the molecular processes that are essential for cell movement¹. It is required for the production of actomyosin-based contractile forces^{2–4}, the regulation of the structure and dynamics of the actin cytoskeleton^{5,6}, and the formation and disassembly of cell–substratum adhesions^{7,8}. Calcium also serves as a second messenger in many biochemical signal-transduction pathways⁷. However, despite the pivotal role of calcium in motile processes, it is not clear how calcium regulates overall cell movement. Here we show that transient increases in intracellular calcium, [Ca²⁺]_i, during the locomotion of fish epithelial keratocytes, occur more frequently in cells that become temporarily ‘stuck’ to the substratum or when subjected to mechanical stretching. We find that calcium transients arise from the activation of stretch-activated calcium channels, which triggers an influx of extracellular calcium. In addition, the subsequent increase in [Ca²⁺]_i is involved in detachment of the rear cell margin. Thus, we have defined a mechanism by which cells can detect and transduce mechanical forces into biochemical signals that can modulate locomotion.

The simple shape and rapid locomotion of fish keratocytes makes them an excellent system in which to study cell movement. These features have been exploited in several studies that have provided insight into how cytoskeletal dynamics, adhesion formation and force generation are integrated to produce movement^{9–15}. Most of this work has focused on fan-shaped keratocytes that exhibit a rapid ‘gliding’ mode of movement. However, these cells can also exhibit a much slower mode of movement closely resembling that of fibroblasts, in which extension of the front edge of the cell occurs as a distinct phase from retraction of the rear edge. Individual keratocytes can also switch between these two modes of movement, which is a particularly useful characteristic for studying how cell movement is regulated.

To investigate the role of calcium in the regulation of keratocyte locomotion, we used fluorescence video microscopy to continuously monitor [Ca²⁺]_i in moving keratocytes (Fig. 1a, b). The resting level of [Ca²⁺]_i was found to be ~138 nM. However,

[†] Present address: Department of Molecular and Cell Biology, University of Connecticut, 75 North Eagleville Road, U-125, Storrs, Connecticut 06269, USA.

keratocytes exhibited transient increases in $[Ca^{2+}]_i$ (Fig. 1c–h) ranging from 219 to 683 nM, corresponding to an increase of between 160 and 500% in fluorescence intensity of the calcium indicator above the baseline value (Fig. 1i). Dual fluorescence imaging of a calcium-insensitive rhodamine dextran and Calcium

Green I dextran that were previously co-loaded into keratocytes confirmed that fluctuations in cell thickness were not contributing to changes in brightness of the calcium indicator (Fig. 1j). The frequency, duration and relative magnitude of calcium transients was compared between fan-shaped cells that exhibited a rapid,

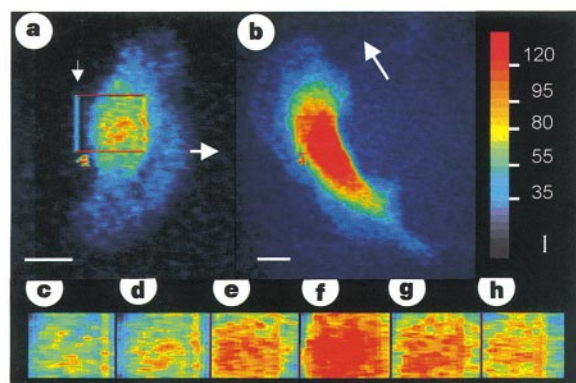
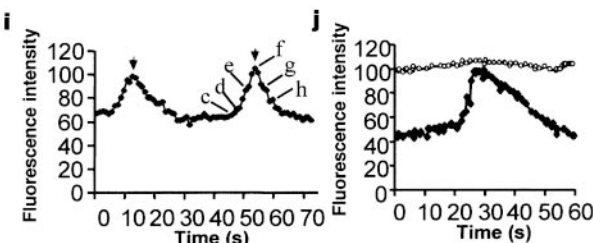


Figure 1 Calcium transients occur in moving keratocytes. **a, b**, Pseudocoloured (non-ratiometric) fluorescence images of a fan-shaped keratocyte (**a**) and a fibroblastic-shaped keratocyte (**b**), moving as indicated (arrows). The colour key represents fluorescence intensity (I) in grey level units. **c–h**, A series of square regions taken every 4 s from the cell in **a** showing a transient increase in $[Ca^{2+}]_i$ as a brief increase in red colour. **i**, Plot of average fluorescence intensity over time within the same region with points corresponding to regions (**c–h**) marked (on the



second transient, arrow). The same relative increase in fluorescence occurs throughout the cell, but the nuclear region appears brighter because it is thicker. However, differences in cell thickness do not affect the observation of transients. **j**, Plot of the fluorescence intensity from a calcium-insensitive rhodamine dextran (circles) shows no change compared with the transient increase in fluorescence from the calcium indicator (squares). Scale bars, 10 μ m.

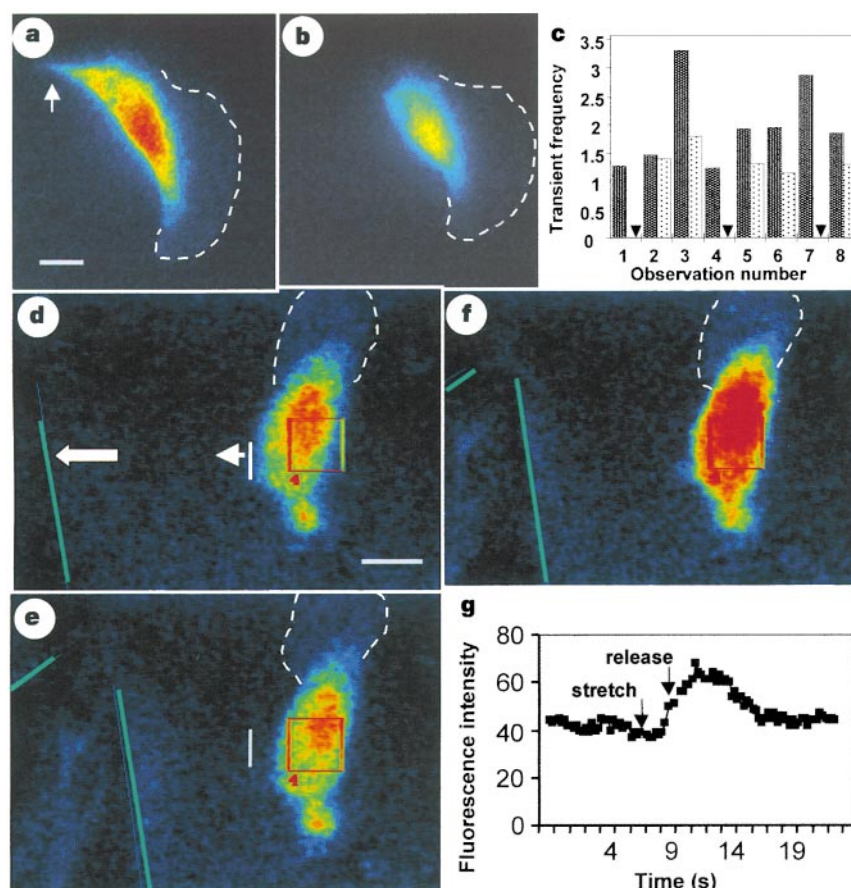


Figure 2 Increases in $[Ca^{2+}]_i$ in response to stretching. **a**, A stretched keratocyte with a 'tether' (arrow). **b**, The same cell following detachment (dashed line represents front cell edge). **c**, Transient frequency (per minute) or zero transients (triangles) in tethered cells (dark bars) and after detachment (pale bars). **d**, A microneedle (green line) is used to deform an elastic substratum in the direction indicated (arrow). The cell is stretched for ~2 s in the same direction (white

vertical bar). **e**, The microneedle and the cell return to their pre-stretch positions. Note the position of the cell body with respect to its stretched position (white bar). **f**, The calcium transient induced in response to stretching is shown as an increase in red over the cell body. **g**, Plot of fluorescence intensity versus time, showing that a calcium transient is induced about 1 s after stretch initiation. Scale bar, 10 μ m.

gliding mode of locomotion and fibroblastic-shaped cells that exhibited a slow, discontinuous mode of movement. Calcium transients were found to occur about three times more frequently in fibroblastic-shaped keratocytes (mean 1.69 per min, s.e.m. 0.33, $n = 25$) than in fan-shaped cells (mean 0.62 per min, s.e.m. 0.22,

$n = 12$). Most fan-shaped keratocytes did not exhibit any transients within an observation period of 2–3 min, whereas most fibroblastic-shaped cells displayed three transients a minute, and sometimes as many as six. No significant difference was found in the duration or magnitude of the transients displayed by these subgroups of keratocytes.

Calcium transients were abolished in medium containing low calcium concentrations, or by the addition of EGTA to the culture medium, indicating that extracellular calcium is required. Treatment with 50 μM verapamil, a voltage-gated calcium-channel inhibitor, led to a 38% ($n = 8$) irreversible reduction in transient frequency. A larger reduction (57%, $n = 12$) in transient frequency was observed following treatment with 10 μM gadolinium (Gd^{3+}), which is known to block stretch-activated calcium channels, indicating that an influx of extracellular calcium through these channels is involved. Gadolinium treatment also inhibited detachment of the rear cell margin, causing cells to become elongated and eventually to cease movement (data not shown). When cells were placed in a high-potassium buffer, the frequency of calcium transients was not increased, as might be expected if voltage-dependent channels were having a significant effect. Treatment of keratocytes with thapsigargin (1 μM), to deplete calcium from intracellular stores, completely inhibited calcium transients, indicating that these arise from a calcium-induced calcium release within the cell. Furthermore, the frequency of calcium transients was markedly reduced (83%, $n = 8$) by treatment with cytochalasin D (250 ng), consistent with a requirement for an intact actin cytoskeleton.

To investigate the role of calcium transients in locomotion, alterations in cell shape were monitored in relation to changes in $[\text{Ca}^{2+}]_i$. In fibroblastic-shaped cells (Fig. 2a), a clear increase in transient frequency was observed as cells become elongated, before the rear cell edge was retracted (Fig. 2b). This is particularly striking for 'tethered' cells, whose movement becomes impeded when a small portion of the cell margin is temporarily unable to detach from the substratum (Fig. 2a). The frequency of transients in tethered cells is greatly reduced following retraction (Fig. 2c). These results, together with the finding that transients could be inhibited by Gd^{3+} , point to the involvement of stretch-activated calcium channels.

To test this idea, experiments were performed to determine whether an externally applied mechanical force could induce calcium transients, by stretching keratocytes without direct physical contact. Cells preloaded with calcium indicator were plated onto elastic substrata and stretched using paired microneedles on either side of the cell to deform the underlying substratum (Fig. 2d–f). In most cases (22 out of 27), a transient increase in $[\text{Ca}^{2+}]_i$ could be induced immediately after stretching (Fig. 2g). Nine of these stretched cells responded to a second stretch stimulus, and only three of these responded to a third stretch, indicating that keratocytes can adapt to repeated stretch stimuli. The increase in fluorescence intensity associated with stretch-induced transients was not significantly different from ones that occur normally.

Direct evidence for the existence of stretch-activated channels in keratocytes was obtained from single-channel recordings within membrane-attached patches ($n = 7$; Fig. 3a) and from whole-cell patch-clamp experiments (Fig. 3b). During the single-channel recording, three consecutive stretching stimuli were given to a small region of the cell membrane by applying negative pressure to the patch electrode. This was followed immediately by a dramatic increase in channel activity (Fig. 3a, top three rows). The spontaneous channel activity is greater at a pipette voltage of 0 mV (horizontal bar) than at +50 mV (remainder of recording). This is consistent with previous observations that stretch-activated channel opening is more frequent with membrane depolarization¹⁶. In addition, the open probability (P_o) increases dramatically during application of negative pressure (bars 1, 2, 3) and decreases rapidly after re-establishing normal pressure, indicating that the recordings

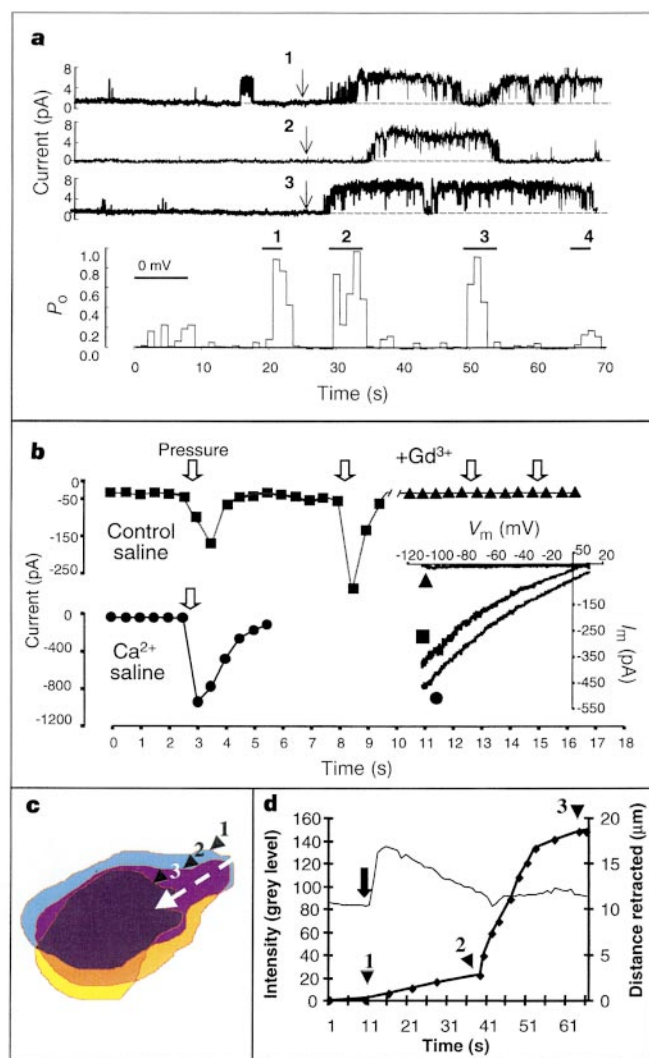


Figure 3 Direct evidence for stretch-activated calcium channels. **a**, Activity of a single stretch-activated channel plotted (traces) for three recording periods of 4 s each centred on the application of negative pressure (arrows 1, 2, 3). Activity of the channels increases markedly each time. The graph (bottom) is a continuous display of the open probability (P_o) for this channel binned into 1 s epochs. Bars 1, 2, 3 and 4 represent the duration of suction. **b**, Stretch-activated channels in keratocytes conduct Ca^{2+} and are blocked by Gd^{3+} . Recordings from stretch-activated channels in the whole-cell patch-clamp configuration, in control saline (squares), Ca^{2+} saline (circles) and following addition of Gd^{3+} (triangles). Deflections in these graphs (open arrows) correspond to the activation of stretch-activated calcium channels after a brief 'puff' of pressure and represent an influx of cations (squares) or calcium (circles). Channel activity in control saline is completely blocked by 100 μM Gd^{3+} (triangles). Representative leak-subtracted channel current-voltage relations (inset) for each condition measured using voltage ramps during the application of pressure. **c**, **d**, Rear retraction induced by the photorelease of caged ionophore. **c**, Cell outlines (triangles 1–3) during retraction (along dashed line). **d**, A plot of calcium indicator fluorescence (plain line) and retraction rate (line with symbols; triangles correspond to outlines in **c**) shows that retraction is initiated (triangle 1) immediately after photoactivation (bold arrow) and the resultant increase in $[\text{Ca}^{2+}]_i$. Retraction proceeds slowly while $[\text{Ca}^{2+}]_i$ returns to its resting level and then proceeds rapidly to completion (triangles 2 and 3).

are channel responses and not nonspecific behaviour, such as seal breakdown (Fig. 3a, bottom graph). Stretch activation could be induced consecutively three times with repeated applications of suction. Furthermore, this channel undergoes an adaptive response to the applied force. This was initially manifest by a decline in P_0 despite continued suction (response 3), and finally by a markedly weakened response to a subsequent stimulus (response 4). Such adaptation of stretch-activated channel responses has previously been reported^{17,18}, and is consistent with reduced cellular responses to repeated stretches, as mentioned above. Finally, despite uncertainty of the absolute value of patch membrane potential (cell interior resting potential was not measured), we estimated the single-channel conductance from the difference in current amplitudes at pipette voltages of 0 and +50 mV. Calculations for four membrane patches yielded values of 51, 52, 55 and 56 pS, within the range of values previously measured for cation-selective stretch-activated channels¹⁹.

Whole-cell patch experiments showed that stretch-activated channels (Fig. 3b) could be activated in response to the application of brief 'puffs' of pressure through the patch pipette. Measurement of stretch-activated channel currents in control extracellular saline were characterized by a current-voltage relation with slight inward rectification, a reversal potential (E_{rev}) of 5 ± 1 mV ($n = 7$) and a complete block by $100 \mu\text{M}$ Gd^{3+} . The activation of stretch-activated channels was shown to mediate an influx of extracellular calcium.

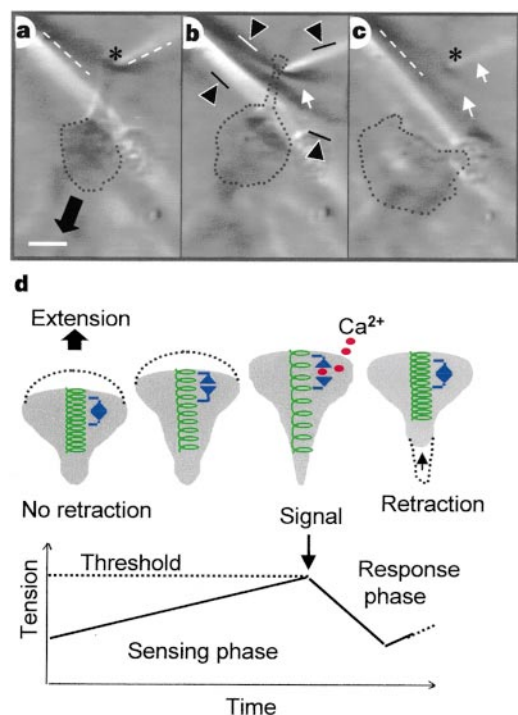


Figure 4 A scheme for the mechano-chemical regulation of movement. **a-c**, Phase-contrast images taken at 5-s intervals of a keratocyte on an elastic substratum. **a**, The front edge (dark dotted line) is extending (black arrow) but the rear (asterisk) is stuck. **b**, Increasing tension compresses the substratum between the regions indicated between opposing arrowheads, causing a new wrinkle to form (white arrow) and the pre-existing wrinkles (white dashed lines in **a**) to become more prominent. **c**, When retraction occurs (from the asterisk), tension is released so that two wrinkles disappear (white arrows) and the remaining wrinkle (white, dashed line) becomes less prominent. The cell now becomes fan shaped (dark dotted line). Scale bar, 10 μm . **d**, A tethered cell outline (grey) showing the cytoskeleton (green spiral) mechanically coupled to stretch-activated calcium channels (blue). During extension (dotted line), the cell is stretched and tension increases (graph). When a critical threshold is reached (dotted line), the calcium channels are activated, causing an influx of extracellular calcium (red) and a transient increase in $[\text{Ca}^{2+}]_i$, which induces rear retraction.

Thus, when Na^+ and K^+ in the extracellular saline were replaced with Ca^{2+} , the stretch-activated channel carried currents of the same density as the control (control, 67 ± 25 pA pF⁻¹, $n = 7$; Ca^{2+} , 61 ± 11 pA pF⁻¹, $n = 5$), demonstrating that these channels are capable of generating a significant intracellular Ca^{2+} signal. The reversal potential of these Ca^{2+} currents was shifted by +25 mV compared to control currents (E_{rev} , 30 ± 6 mV; $n = 5$), indicating that the channel is slightly more permeable to Ca^{2+} than to Na^+ , when both are compared to Cs^+ , which was the main internal cation in these experiments.

To investigate the role of increased $[\text{Ca}^{2+}]_i$ in the regulation of keratocyte movement, photoactivation of caged compounds was used to induce calcium transients, and the resulting alterations in cell shape were observed (Fig. 3c, d). Induction of a calcium transient could be achieved after the photorelease of either the caged ionophore A23187, which triggers an influx of extracellular calcium, or inositol 1,4,5-triphosphate, which induces Ca^{2+} release from intracellular stores. In both cases, induction of a calcium transient was usually followed by retraction of the rear cell margin (Fig. 3c). Retraction of the cell margin occurred preferentially at the rear of the cell, even though the relative magnitude of the $[\text{Ca}^{2+}]_i$ increase is the same throughout the cell. Retractions were not observed in control experiments in which cells were exposed to ultraviolet in the absence of caged compound.

Our experimental data, together with observations of tethered keratocytes on wrinkling silicone substrata (Fig. 4a-c), suggest the following scheme for the mechanochemical regulation of cell movement. When the rate of retraction becomes much slower than the rate of lamellar extension, the cell will become elongated and tension will increase between each end of the cell (Fig. 3a, b, d), inhibiting lamellar extension²⁰ and reducing overall cell speed. When cytoskeletal tension exceeds a critical threshold, stretch-activated calcium channels are activated, triggering an influx of extracellular calcium and a subsequent release of calcium from intracellular stores, which is typical of other mechanochemical signal-transduction systems²¹⁻²³. The transient increase in $[\text{Ca}^{2+}]_i$ can then activate various detachment mechanisms that lead to retraction of the cell rear. Possible calcium-dependent detachment mechanisms include: an increased actomyosin-based contractile force^{3,24} to pull up the rear cell edge; the disassembly of cell-substratum adhesions by calcium-dependent phosphatases²⁵; and proteases^{26,27} acting independently or in combination with increased cytoskeletal tension^{9,28}. Cytoskeletal tension is decreased after retraction (Fig. 4c, d), removing inhibition of lamellar extension so cell movement can resume.

The activation of stretch-activated calcium channels might account for the spatial²⁸ and temporal²⁹ changes in $[\text{Ca}^{2+}]_i$ observed in other moving cell types. Thus the function of stretch-activated calcium channels could represent a general mechanism by which cell movement is regulated. □

Methods

Cell culture. Keratocytes from molly fish (*Poecilia spheonops*) or goldfish (*Carassius auratus*) were cultured on coverslips as described¹⁴. In stretching experiments, cells were dissociated and replated on elastic substrata as described¹³.

Bead loading. Dextran-conjugated (relative molecular mass, 10,000) calcium indicator Calcium Green I or a calcium-insensitive rhodamine dextran was incorporated into keratocytes as described³⁰.

Calibration of the calcium indicator. The calcium indicator was calibrated using the Calcium Calibration kit (#1) from Molecular Probes. The variability of calcium indicator concentration was assessed by loading Calcium Green I dextran with a calcium-insensitive rhodamine-labelled dextran of comparable molecular size ($M_r = 10,000$). The ratio of Calcium Green I to rhodamine fluorescence was similar in the 20 cells measured, independent of the amount of indicator loaded into the cell, indicating that the resting level of $[\text{Ca}^{2+}]_i$ is similar in different cells. To estimate the range of $[\text{Ca}^{2+}]_i$ increases, calcium

calibration experiments were performed in about 10 individual cells that had been coloaded with both rhodamine and Calcium Green I, following the observation of 1–2 transients.

Calcium imaging. Cell preparation and calcium imaging were performed as described³⁰. Cells attached to coverslips were placed in the well of a metal slide chamber filled with ~0.5 ml of either PBS (with calcium and magnesium) or Fish Ringer solution (containing (in mM): 112 NaCl, 2 KCl, 2.4 NaHCO₃, 1 CaCl₂, 0.5 MgCl₂, 1 TrisCl, pH 7.3. Images were acquired and processed using an intensified CCD camera (MTI 72CCD) in manual mode, and Image 1 software, respectively. Relative changes in fluorescence intensity of the calcium indicator were measured over the nuclear region because the signal-to-noise ratio is higher in this area than in lamellar regions. Camera settings, contrast and brightness settings of Image 1 were the same as for the calibration experiments. Fluorescence images were pseudocoloured for easier observation of calcium transients.

Photoactivation. Photolysis of caged compounds was performed as described³⁰. Caged calcium ionophore was used at a final concentration of ~10 μ M. Caged inositol 1,4,5-trisphosphate was from Calbiochem-Novabiochem (La Jolla), and the caged calcium ionophore A23187 was synthesized by Molecular Probes.

Cell stretching. Elastic substrata were made as described¹³ using the silicone elastomer Sylgard 184 (Dow Corning). Cells previously loaded with calcium indicator were replated onto the elastic substratum. Using a pair of Narashige micromanipulators, glass microneedles were lowered down either side of a cell until they just touched the substratum. With one microneedle held stationary, the other was used to pull the substratum in the opposite direction for 1–2 s. Deformation of cell shape in the direction of the microneedle movement confirmed that the cell had been stretched.

Patch clamping. Cells attached to 12-mm² glass coverslips were placed in a glass-bottomed superfusion chamber containing a solution containing (in mM) 145 NaCl, 5 KCl, 2 CaCl₂, 1 MgCl₂, 10 HEPES, 10 glucose (pH 7.4) situated on the stage of an inverted microscope (Nikon Diaphot). Cells were viewed at a magnification of $\times 400$ using Hoffman modulation contrast optics. Patch pipettes were filled with a standard solution containing (in mM) 150 KCl, 2 MgCl₂, 10 HEPES. Cell-attached membrane patches were isolated by forming gigaohm seals (>20 G Ω) using fire-polished glass (Drummond N51A) pipettes with tip openings of ~1 μ m diameter and electrical resistance of 2–4 M Ω . Currents were measured with an Axopatch 1C amplifier and continuously logged onto a computer hard disk using Axotape software and a Digidata 1200 interface. Single stretch-activated channels were identified by their activation upon application of negative pressure to the patch pipette and subsequent deactivation upon return to atmospheric pressure. Patch recordings usually lasted less than 2 min because of the rapid movement of keratocytes. In the whole-cell patch-clamp configuration, stretch-activated channels were activated by brief puffs of pressure (1–2 s) applied by mouth through the patch pipette, and channel current was monitored by applying a 200-ms voltage ramp from -110 to +10 mV every 500 ms. Currents were recorded with an EPS-9 amplifier (HEKA), digitized and filtered at 5 kHz (4-pole Bessel) and analysed using Pulsefit and Microsoft Excel. Stretch-activated channel current-voltage relations were obtained by subtracting current measured at each potential in the absence of pipette pressure from that recorded with pressure. Patch pipettes with tip diameters of 3–4 μ m (2 M Ω) were pulled from Fisher 50- μ l capillary glass, fire polished and filled with an internal saline containing (in mM) 140 aspartate, 10 EGTA, 1 CaCl₂, 1 MgCl₂, 10 HEPES (pH 7.3 with CsOH). Control

external saline contained 145 NaCl, 5 KCl, 2 CaCl₂, 1 MgCl₂, 10 HEPES (pH 7.4 with NaOH). Calcium saline contained 127 CaCl₂, 1 MgCl₂, 10 HEPES (pH 7.4 with TrisOH). Gd³⁺ (100 μ M) was applied to the bath from a 6X stock in control saline.

Received 16 March; accepted 26 May 1999.

1. Stossel, T. P. On the crawling of animal cells. *Science* **260**, 1086–1094 (1993).
2. Strohmeyer, R. & Bereiter-Hahn, J. Control of cell shape and locomotion by external calcium. *Exp. Cell Res.* **154**, 412–420 (1984).
3. Citi, S. & Kendrick-Jones, J. Regulations of non-muscle myosin structure and function. *BioEssays* **7**, 155–159 (1987).
4. Rees, D. A. *et al.* Myosin regulation and calcium transients in fibroblast shape change: Attachment and patching. *Cell Motil. Cytoskel.* **13**, 112–122 (1989).
5. Condeelis, J. Life at the leading edge: The formation of cell protrusions. *Annu. Rev. Cell Biol.* **9**, 411–444 (1993).
6. Hartwig, J. H. & Yin, H. The organization and regulation of the macrophage actin skeleton. *Cell Motil. Cytoskel.* **10**, 117–125 (1988).
7. Sjaastad, M. D. & Nelson, W. J. Integrin-mediated calcium signaling and regulation of cell adhesion by intracellular calcium. *BioEssays* **19**, 47–55 (1997).
8. Crowley, E. & Horwitz, A. F. Tyrosine phosphorylation and cytoskeletal tension regulate the release of fibroblast adhesions. *J. Cell Biol.* **131**, 525–537 (1995).
9. Lee, J. & Jacobson, K. The composition and dynamics of cell-substratum adhesions in locomoting fish keratocytes. *J. Cell Sci.* **110**, 2833–2844 (1997).
10. Svitkina, T. M., Verkhovskiy, A. B., McQuade, K. M. & Borisov, G. G. Analysis of the actin-myosin II system in fish epidermal keratocytes: Mechanism of cell body translocation. *J. Cell Biol.* **139**, 397–415 (1997).
11. Galbraith, C. G. & Sheetz, M. P. Relationship between cell shape, traction force, and speed. *Mol. Biol. Cell* (suppl.) **8**, 385 (1997).
12. Anderson, K. L., Wang, Y.-L. & Small, J. V. Coordination of protrusion and translocation of the keratocyte involves rolling of the cell body. *J. Cell Biol.* **134**, 1209–1218 (1996).
13. Lee, J., Leonard, M., Oliver, T., Ishihara, A. & Jacobson, K. Traction forces generated by locomoting keratocytes. *J. Cell Biol.* **127**, 1957–1964 (1994).
14. Lee, J., Ishihara, A., Theriot, J. A. & Jacobson, K. Principles of locomotion for simple-shaped cells. *Nature* **362**, 167–171 (1993).
15. Theriot, J. A. & Mitchison, T. J. Actin microfilament dynamics in locomoting cells. *Nature* **352**, 126–131 (1991).
16. Sachs, F. In *Sensory Transduction* (eds Corey, D. P. & Roper, S.) 242–260 (Rockefeller University Press, New York, 1992).
17. Gustin, M. C., Zhou, X., Martinac, B. & Kung, C. A mechanosensitive ion channel in the yeast plasma membrane. *Science* **242**, 762–765 (1988).
18. Hamill, O. P. & McBride, D. W. Rapid adaptation of single mechanosensitive channels in *Xenopus* oocytes. *Proc. Natl Acad. Sci. USA* **89**, 7462–7466 (1992).
19. Sackin, H. In *Molecular Biology of Membrane Transport Disorders* (ed. Schultz, S. G.) 201–222 (Plenum, New York, 1996).
20. Kolega, J. Effects of mechanical tension on protrusive activity and microfilament and intermediate filament organization in an epidermal epithelium moving in culture. *J. Cell Biol.* **102**, 1400–1411 (1986).
21. Pommerenke, H. *et al.* Stimulation of integrin receptors using a magnetic drag force device induces an intracellular free calcium response. *Eur. J. Cell Biol.* **70**, 157–164 (1996).
22. Glogauer, M., Ferrier, J. & McCulloch, C. A. G. Magnetic fields applied to collagen-coated ferric oxide beads induce stretch-activated Ca²⁺ flux in human fibroblasts. *Am. J. Physiol.* **269**, C1093–C1104 (1995).
23. Naruse, K. & Sokabe, M. Involvement of stretch-activated ion channels in Ca²⁺ mobilization to mechanical stretch in endothelial cells. *Am. J. Physiol.* **264**, 1037–1044 (1993).
24. Chen, W.-T. Mechanism of retraction of the trailing edge during fibroblast movement. *J. Cell Biol.* **90**, 187–200 (1981).
25. Hendey, B., Klee, C. B. & Maxfield, F. R. Inhibition of neutrophil chemokinesis on vitronectin by inhibitors of calcineurin. *Science* **258**, 296–299 (1992).
26. Huttenlocher, A. *et al.* Regulation of cell migration by the calcium-dependent protease calpain. *J. Biol. Chem.* **272**, 32719–32722 (1997).
27. Palecek, S. P., Huttenlocher, A., Horwitz, A. F. & Lauffenburger, D. A. Physical and biochemical regulation of integrin release during rear detachment of migrating cells. *J. Cell Sci.* **111**, 929–940 (1998).
28. Brundage, R. A., Fogarty, K. E., Tuft, R. A. & Fay, F. S. Calcium gradients underlying polarization and chemotaxis of eosinophils. *Science* **254**, 703–706 (1991).
29. Marks, P. W. & Maxfield, F. R. Transient increases in cytosolic free calcium appear to be required for the migration of adherent human neutrophils. *J. Cell Biol.* **110**, 43–52 (1990).
30. Ishihara, A., Gee, K., Schwartz, S., Jacobson, K. & Lee, J. Photoactivation of caged compounds in single, living cells: An application to the study of cell locomotion. *Biotechniques* **23**, 268–274 (1997).

Acknowledgements. This work was supported by the NIH (K.J. and G.O.).

Correspondence and requests for materials should be addressed to J.L. (e-mail: jlee@uconnvm.uconn.edu).